

GenCore version 4.5
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: May 17, 2001, 12:27:29 ; Search time 1258.03 Seconds
(without alignments)
11078.215 Million cell updates/sec

Title: US-09-462-846-1
Perfect score: 945
Sequence: 1 atgcagctgaacctatt.....aatgatcgtctccatctg 945

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 1283235 seqs, 7373929652 residues
Total number of hits satisfying chosen parameters: 2566470

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : GenEmbl:
1: gb_ba1:*
2: gb_ba2:*
3: gb_ba3:*
4: gb_in1:*
5: gb_in2:*
6: gb_in3:*
7: gb_om:*
8: gb_om:*
9: gb_ov:*
10: gb_pat2:*
11: gb_ph:*
12: gb_pi1:*
13: gb_pi2:*
14: gb_pi3:*
15: gb_pi4:*
16: em_ba1:*
17: em_ba2:*
18: em_fun:*
19: em_hrgo_hum:*
20: em_hrgo_inv:*
21: em_hrgo_rod:*
22: em_hrgo_hum1:*
23: em_hrgo_hum2:*
24: em_hrgo_hum3:*
25: em_hrgo_hum4:*
26: em_hrgo_hum5:*
27: em_hrgo_hum6:*
28: em_hrgo_hum7:*
29: em_hrgo_hum8:*
30: em_hrgo_inv1:*
31: em_hrgo_inv2:*
32: em_hrgo_other:*
33: em_hrgo_rod:*
34: em_hum1:*
35: em_hum2:*
36: em_hum3:*
37: em_hum4:*
38: em_hum5:*
39: em_hum6:*
40: em_hum7:*
41: em_in:*
42: em_om:*
43: em_or:*

44: em_ov:*
45: em_pat:*
46: em_ph:*
47: em_pi:*
48: em_ro:*
49: em_sts:*
50: em_sy:*
51: em_un:*
52: em_v1:*
53: gb_sts1:*
54: gb_sts2:*
55: gb_sts3:*
56: gb_sy:*
57: gb_un:*
58: gb_v11:*
59: gb_v12:*
60: gb_hrg1:*
61: gb_hrg2:*
62: gb_hrg3:*
63: gb_hrg4:*
64: gb_hrg5:*
65: gb_hrg6:*
66: gb_hrg7:*
67: gb_hrg8:*
68: gb_hrg9:*
69: gb_hrg10:*
70: gb_hrg11:*
71: gb_hrg12:*
72: gb_hrg13:*
73: gb_hrg14:*
74: gb_hrg15:*
75: gb_hrg16:*
76: gb_hrg17:*
77: gb_hrg18:*
78: gb_hrg19:*
79: gb_hrg20:*
80: gb_hrg21:*
81: gb_hrg22:*
82: gb_hrg23:*
83: gb_hrg24:*
84: gb_hrg25:*
85: gb_pi1:*
86: gb_pi2:*
87: gb_pi3:*
88: gb_pi4:*
89: gb_pi5:*
90: gb_pi6:*
91: gb_pi7:*
92: gb_pi8:*
93: gb_pi9:*
94: gb_pi10:*
95: gb_pi11:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Match	Length	DB	ID	Description
1	945	100.0	945	9	AX006200	AX006200 Sequence
2	945	100.0	216750	2	BSDB0007	Z99110 Bacillus su
3	320.4	33.9	36308	2	D88802	D88802 Bacillus su
4	320.4	33.9	213190	2	BSDB0004	Z99107 Bacillus su
5	319	33.8	945	9	AX006205	AX006205 Sequence
6	311.8	33.0	948	2	AX006203	AX006203 Sequence
7	311.8	33.0	3962	2	BSDB02562	U02562 Bacillus su
8	311.8	33.0	4162	2	BACORFX	D45048 Bacillus su
9	311.8	33.0	212610	2	BSDB0019	Z99122 Bacillus su
10	304	32.2	15012	1	AB011838	AB011838 Bacillus
11	304	32.2	305153	2	AP001520	AP001520 Bacillus

```
12 277.2 29.3 10607 1 AE006310 AE006310 Lactococc
13 265.8 28.1 1379 3 STREPM D16594 S. mutans pm
14 252.2 26.7 49897 71 AC027136 AC027136 Staphyloc
15 227 24.0 2815 AF269779 AF269779 Staphyloc
16 226 23.9 147006 84 SPNEU1904 AL449926 Streptoco
17 225.8 23.9 2111 A64844 A64844 sequence 1
18 192.4 20.4 44449 76 AC027830 AC027830 Staphyloc
19 192.4 20.4 49487 75 AC074320 AC074320 Staphyloc
20 192.4 20.4 55389 1 AC025592 AC025592 Staphyloc
21 192.4 20.4 55389 1 AC078934 AC078934 Staphyloc
22 189.4 20.0 3473 AF269301 AF269301 Staphyloc
23 184.4 14.9 2423 AF270310 AF270310 Staphyloc
24 138.4 14.6 3754 AB016431 AB016431 Staphyloc
25 59.2 6.3 3883 2 LLAU5118 AU005118 Lactococc
26 58.2 6.2 619 AF015751 AF015751 Lactococc
27 57 6.0 213732 1 AE001862 AE001862 Deinococc
28 56.4 6.0 21343 1 AE001744 AE001744 Thermocog
29 53.8 5.7 6951 1 AF001955 AF001955 Streptoco
30 50 5.3 7218 10 AC004157 AC004157 Sequence 14
31 43.4 4.6 169546 60 AC004157 AC004157 Plasmodi
32 39.4 4.2 94239 13 AF014P22 AL137082 Arabidops
33 38.6 4.1 87835 12 AB005237 AB005237 Arabidops
34 38.2 4.0 161528 74 AC005435 AC005435 Homo sap1
35 38.2 4.0 172064 90 AL445195 AL445195 Human DNA
36 38.2 4.0 175012 74 AC069072 AC069072 Homo sap1
37 38.2 4.0 193378 64 AC01591 AC01591 Homo sap1
38 38.2 4.0 193858 65 AC016760 AC016760 Homo sap1
39 38 4.0 2825 6 THESPAGI M63017 Theileria a
40 38 4.0 168778 92 HSA490024 AL117338 Human DNA
41 38 4.0 189256 81 AL391004 AL391004 Homo sap1
42 37.8 4.0 140064 13 ATAC009177 AC009177 Arabidops
43 37.6 4.0 195721 71 AC027059 AC027059 Homo sap1
44 37.4 4.0 85356 13 ATAC49 AL080318 Arabidops
45 37.4 4.0 99814 13 ATR8J2 AL132269 Arabidops
```

ALIGNMENTS

```
RESULT 1
LOCUS AX006200 945 bp DNA PAT 24-NOV-2000
DEFINITION Sequence 1 from Patent WO9504016.
ACCESSION AX006200
VERSION AX006200.1 GI:9929071
KEYWORDS
SOURCE
ORGANISM Bacillus subtilis.
Bacteria; Firmicutes; Bacillus/Clostridium group; Bacillus.
REFERENCE 1 (bases 1 to 945)
Bacillus/Staphylococcus group; Bacillus.
AUTHORS Estell, D.A.
JOURNAL Patent: WO 9504016-A, 1 28-JAN-1999;
GENENOR INTERNATIONAL P.V (NL); ESTELL DAVID A (US)
FEATURES
Source 1..945
/organism="Bacillus subtilis"
/db_xref="taxon:1423"
BASE COUNT 288 a 180 c 232 g 245 t
RIGIN
```

```
Query Match 100.0%; Score 945; DB 9; Length 945;
Best Local Similarity 100.0%; Pred. No. 1.2e-250;
Matches 945; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
```

```
Y 1 atgacgactgaacgctgattttcaagcgtgtttcaagaagaattggggcggagac 60
b 1 ATGACGACTGAACCGTATTTTTCAAGCGTGTTCACAAAGAAATTGGGGCGGAGACC 60
Y 61 gcttaagctgatttggcctatcacattccgtcacaagcaagaggagtgctggccttt 120
|||||
```

```
D6 61 GCTTTAGCTGATTTTGGCTATACCATTTCCGTCACACGACGAGGAGTGGCTTTT 120
QY 121 gccgcgcatcaaaatggtcaaaagcgtgtgttcaaaaacggaatgtatcaagggttcacgctc 180
D6 121 GCCGCGCATCAAAATGTCAAAGCGTGTTCAAAACGGAATGTATTAAGGGGTACGCTC 180
QY 181 agcgaattatgggaacatcacacacacattatctcgacacagcttgaaaggagacgtttccct 240
D6 181 AGCGAATTAAGGGAACATCACACACATTATTCGACACAGCTTGAAGGGGACCGTTCCCT 240
QY 241 ctgcttaacaaaatattagatgtctgaacgaactatctgttcagtgatccgaagat 300
D6 241 CTGCTTAACAAAATATTAGATGCTGACGAGACTATTCGTTCAGTGCATCCGATAT 300
QY 301 gaatatgcacacatacatgaacacggttgagcttggaacacgaatgtgtgtacattt 360
D6 301 GAATATGCCAACATACATGATGAACGCTGAGCTTGGAACACGAATGCTGTGATATT 360
QY 361 gattgccaacaaagatccgagattatttatttgccacaacatgcaacaaagaagaacta 420
D6 361 GATTGCCAACAAAGATCCGAGATTATTATTATGCGCAACATGCAACAAAGAGAACTA 420
QY 421 actacacatgataagcgttgagatgagatgagatgagctcttgccgctgttaaggttaagcgc 480
D6 421 ACTACCATGATGAGCGTGTGAGATGAGATGAGATGAGCTTTCGCCGCTGTAAAGTTAAGCGC 480
QY 481 ggggatttttctatgtgcacagcggttactgttcacatgacgattggaagaattcttgc 540
D6 481 GGGGATTTTCTATGTGTGCGCACGCGTACTGTTCATGCGATGGAAGAAAGAAATTTCT 540
QY 541 ttggaagcgcagacagaaactcaacacacacacacacacacacacacacacacacacacacac 600
D6 541 TTGGAAGCGCAGCAGAACTCAGACACACACACACACACACACACACACACACACACACAC 600
QY 601 gcaagaagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagc 660
D6 601 GCAGAAAGCAGCAGCTGCGCAGCAGCTTCATCTGAAGAAAGCAGTGAAGTATAGAGTCCCG 660
QY 721 tctatccagaagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagc 720
D6 721 TCTATTCACAAAGCGCATACGTCACATGACATGACATGACATGACATGACATGACATGAC 720
QY 781 ttgatgaatgcgcttacttctcggttggaagaaatggaacttaccagcagcaagctta 780
D6 781 TTGATTAATGCGCTTACTTCTCGGTGGGAAATGCAACTTATCAGGATTCACCAAGCTTA 780
QY 841 aagcagaacaaacacatcccttcttcatcagtgatgaagggagggcgatgatctct 840
D6 841 AAGCAGCAAAACCAATCCCTTCTTATCAGTGAAGGAGGCGCGATGATCTCT 840
QY 841 ggtgagatgtctatccttcaaaaagagatcatatgttgcctccttaaggtctga 900
D6 841 GGTGAGATGTCTATCCTTCAAAAAGAGATCATATGTTGCTGCTTACGCTTTCGA 900
QY 901 gaattaaactcgaagatattcagaatgtatcgtctcccatctg 945
D6 901 GAATTAACTCGAAGATATGCAATGATGCTCCCATCTG 945
```

```
RESULT 2
LOCUS BSUB0007 216750 bp DNA BCT 26-NOV-1997
DEFINITION Bacillus subtilis complete genome (section 7 of 21): from 1194391
ACCESSION 299110 AL009126
VERSION 299110.1 GI:2633472
KEYWORDS
SOURCE Bacillus subtilis.
ORGANISM Bacillus subtilis.
Bacteria; Firmicutes; Bacillus/Clostridium group;
Bacillus/Staphylococcus group; Bacillus.
REFERENCE 1 (bases 1 to 216750)
Kunst, F., Ogasawara, N., Moszer, I., Albertini, A.M., Alloni, G.,
```

Azevedo,V., Bertero,M.G., Bessieres,P., Bolotin,A., Borchert,S., Boriss,R., Boursier,L., Brans,A., Braun,M., Brignell,S.C., Bron,S., Brouillet,S., Brusch,C.V., Caldwell,B., Capuano,V., Carter,N.M., Choi,S.K., Codani,J.J., Conerton,I.F., Cummings,N.J., Daniel,R.A., Denzot,F., Devine,K.M., Dusterhoff,A., Ehrlich,S.D., Emerson,P.T., Entlan,K.D., Errington,J., Fabret,C., Ferrari,E., Foulger,D., Fritz,C., Fujita,M., Fujita,Y., Fuma,S., Galizzi,A., Galleron,N., Ghim,S.Y., Glaser,P., Goffeau,A., Golightly,E.J., Grallion,G., Guiseppi,G., Guy,B.J., Haga,K., Hacheh,J., Harwood,C.R., Henaut,A., Hilbert,H., Holsappel,S., Hosono,S., Hultoy,M.F., Itaya,M., Jones,L., Joris,B., Karamata,D., Kasahara,Y., Klauer,Blanchard,M., Klein,C., Kobayashi,T., Koelter,P., Koningstein,G., Krogh,S., Kumano,M., Kurita,K., Lapidus,A., Lardinot,S., Maubert,J., Lazarevic,V., Lee,S.M., Levine,A., Liu,H., Masuda,S., Mauel,C., Medigue,C., Medina,N., Mellado,R.P., Mizuno,M., Moestl,D., Nakai,S., Noback,M., Noone,D., O'Reilly,M., Ogawa,K., Ogilwara,A., Oudega,B., Park,S.H., Parro,V., Pohl,T.M., Portelle,D., Portollik,S., Prescott,A.M., Presecan,E., Pujic,P., Purnelle,B., Rapoport,G., Rey,M., Reynolds,S., Rieger,M., Rivolta,C., Roche,E., Roche,B., Rose,M., Sadale,Y., Sato,T., Scallan,E., Schleich,S., Schroeter,R., Scoffone,P., Sekiguchi,J., Sekowska,A., Seror,S.J., Serrero,P., Shin,B.S., Solio,B., Sorokin,A., Tacconi,E., Takagi,T., Takahashi,H., Takemaru,K., Takeuchi,M., Tamakoshi,A., Tanaka,T., Terpeita,P., Tognoni,A., Tostato,V., Uchiyama,S., Vandenbol,M., Vanher,P., Vassarotti,A., Viari,A., Wambuit,R., Wedler,E., Wedler,H., Weltenegeger,T., Winters,P., Wipal,A., Yamamoto,H., Yamane,K., Yasumoto,K., Yata,K., Yoshida,K., Yoshikawa,H.F., Zumstein,E., Yoshikawa,H. and Danchin,A.

The complete genome sequence of the gram-positive bacterium *Bacillus subtilis* Nature 390 (6657), 249-256 (1997) 98044033

2 (bases 1 to 216750)

Kunst,F., Ogasawara,N., Yoshikawa,H. and Danchin,A.

Direct Submission

Submitted (18-NOV-1997) I. Moszer, A. Danchin, Institut Pasteur, Regulation de l'Expression Genetique, 28 rue du Docteur Roux, 75724 Paris Cedex 15, FRANCE. E-mail: moszer@pasteur.fr, adanchin@pasteur.fr Phone: +33 (0)1 45 68 84 41, Fax: +33 (0)1 45 68 89 48

FEATURES

source

1. 216750

/organism="Bacillus subtilis"

/strain="168"

/db_xref="taxon:1423"

116..1156

/gene="argC"

116..1156

/gene="argC"

/EC_number="1.2.1.38"

/function="arginine biosynthesis"

/codon_start=1

/transl_table=11

/product="N-acetylglutamate gamma-semialdehyde dehydrogenase"

/protein_id="CAB12976.1"

/db_xref="GI:2633473"

/translation="MKIGVATGCGTELVRIIISHHPHAECTIYSSSGEGNVYSEGVYHITGADLOOLKPTDMNTIKHEIDIMFLAPPGVSSSLPKIADAGITVYDLSGLR IKEPAYEKYKRTAPKPAVIOEAUYGIAENLOIOAKIINAGCGFPYAVLIGLAP LAKKILDSFVIYDAKTGVSAGAKASMGTHFSELNDNRITRYVNHQHTPELBOAT NEMQEPOLGPTSAHLVFNTRIGIMKIKTRITICDILADLDLISFETQSYRYRAP KQOYPOTKREVSNFCDIAVLDERTNRTVIVSIDNLKGAQOAVONFNLVANGME EIGLTIPIYP"

1176..2396

/gene="argJ"

1176..2396

/gene="argJ"

/EC_number="2.3.1.35"

/EC_number="2.3.1.1"

/function="arginine biosynthesis"

gene

CDS

1176..2396

/gene="argJ"

1176..2396

/gene="argJ"

/EC_number="2.3.1.35"

/EC_number="2.3.1.1"

/function="arginine biosynthesis"

gene

CDS

3184..4341

/gene="argD"

3184..4341

/gene="argD"

/EC_number="2.6.1.11"

/function="arginine biosynthesis"

/codon_start=1

/transl_table=11

/product="N-acetylornithine aminotransferase"

/protein_id="CAB12979.1"

/db_xref="GI:2633476"

/translation="MSLFTQYGRWDIDIKKAKGYVEDONKTYLDFIOGIAVSNLGHCEAVTEAVKQLODSYWHYSNLFQNSLOQAOKIAGSAGDLYFPNGSGOANEGA YKIAARKAKGTCKIITFLQSGPHRTAGMATGODIKTGFPGMLGFFLYLPNDPAF KALCEBGDIAXVMLFTYQSGGVNPAASAEPLSAVOSFOCKOALLIDEIOTGIRG KGFAYEHRGSPDITTYAKGIGKGFPGVYGVICKOIGEPATPGSGTFRGGMILAMA VNATLOIVFQPDLEKADGAPLAKFQDLAEKSPYKDIRKGMIGTIBDDGYADI IAELOTLGLVLPAGPERDSAAPLTVTDEIAEAVSKLOALAHNSAVNO"

4412..5473

/gene="carA"

4412..5473

/gene="carA"

/EC_number="6.3.5.5"

/function="arginine biosynthesis"

/note="alternate gene name: cpa"

/codon_start=1

/transl_table=11

/product="carbamoyl-phosphate transferase-arginine (subunit A)"

/protein_id="CAB12980.1"

/db_xref="GI:2633477"

/translation="WSTLVLEDGTSFSGELDQHENCTGEAVFTGNTGYOEVLDPS YKGOIVFTYPLNGTINENKDFESKRPQVKAAYVTEADQHSRIEAYVSLSEIOLKQ NIPILTVDRVAVVKIRANGTGAATVTSKBSAEATLADPEVAQASQAEISTFGG MKNIALIDFGKKSIASSLVKRGCKVTVVYQDMEEVYIKPDGIVLSNGPDEPAIO PYLGRKISISRPPTLIGICGHOLIALAGGKFMKIPFGHRANMPVINDRTKRVEMT SONSHVYDEQSIINEELTTRFHHVNDTSVEGLAHKKLIPVMSVQFHPRAHQAASEM IFDDYLNKVIIPAREIAHA"

/note="alternate gene name: argA, argE; ornithine acetyltransferase"

/codon_start=1

/transl_table=11

/product="amino-acid acetyltransferase"

/protein_id="CAB12977.1"

/db_xref="GI:2633474"

/translation="MTQISEQIVKTVGVSPKGFQAKGVHCGIARYSKKDIAYIIE TPVSAAYVQSHFOAAPIKYTDQSLKSGPRLKATVYSAIANACTBOGICAKAYTMR ESPASOLGTEPELVAVSSTGVYGEHLDMKIHAGIELIKPPAGSGDEEELITTDVY IKGTETELAGKGTVTYTGARKSGVIRPMAITMGFTPTTAAIEKALQKALRETD VSNQUTVGEISTNDVYVMAVNAENECLEDDPVPKRALCLCEDELEIKR DGGARLEAEAOVGAKNNLDNAVIAKKTIVSNLVKTAIVYOTDNWGMTIIGAIGHSA OVAEEVLEYLGQCLFKNNRQPPESIAKRYLEGGDEITVIMARQDNGRAMGCD LTYDYIKINASYPT"

2411..3187

/gene="argB"

2411..3187

/gene="argB"

/EC_number="2.7.2.8"

/function="arginine biosynthesis"

/codon_start=1

/transl_table=11

/product="N-acetylglutamate 5-phosphotransferase"

/protein_id="CAB12978.1"

/db_xref="GI:2633475"

/translation="MKRTIVFGCGSVIIEISEEYFONIKELRASGMKLAIVHGQPE IYTNMLKRLNKTFFESGQRTTKPVLEVAEWLSSGVKFFVAELAKGLAAAGISK DGLLEADYDPEPTGEVGEIKKVDASVNAIEMNGIIPVLAIPMSISDCTLVVND LAASVAGALEADKLMEFTVDVGIKMKRQRLDVLPRKTIOMIKOEVITGMIPKVS ALSLSDPQSEVMIVNGKSFPAEQTFQGTIKVKAKEAVS"

3184..4341

/gene="argD"

3184..4341

/gene="argD"

/EC_number="2.6.1.11"

/function="arginine biosynthesis"

/codon_start=1

/transl_table=11

/product="N-acetylornithine aminotransferase"

/protein_id="CAB12979.1"

/db_xref="GI:2633476"

/translation="MSLFTQYGRWDIDIKKAKGYVEDONKTYLDFIOGIAVSNLGHCEAVTEAVKQLODSYWHYSNLFQNSLOQAOKIAGSAGDLYFPNGSGOANEGA YKIAARKAKGTCKIITFLQSGPHRTAGMATGODIKTGFPGMLGFFLYLPNDPAF KALCEBGDIAXVMLFTYQSGGVNPAASAEPLSAVOSFOCKOALLIDEIOTGIRG KGFAYEHRGSPDITTYAKGIGKGFPGVYGVICKOIGEPATPGSGTFRGGMILAMA VNATLOIVFQPDLEKADGAPLAKFQDLAEKSPYKDIRKGMIGTIBDDGYADI IAELOTLGLVLPAGPERDSAAPLTVTDEIAEAVSKLOALAHNSAVNO"

4412..5473

/gene="carA"

4412..5473

/gene="carA"

/EC_number="6.3.5.5"

/function="arginine biosynthesis"

/note="alternate gene name: cpa"

/codon_start=1

/transl_table=11

/product="carbamoyl-phosphate transferase-arginine (subunit A)"

/protein_id="CAB12980.1"

/db_xref="GI:2633477"

/translation="WSTLVLEDGTSFSGELDQHENCTGEAVFTGNTGYOEVLDPS YKGOIVFTYPLNGTINENKDFESKRPQVKAAYVTEADQHSRIEAYVSLSEIOLKQ NIPILTVDRVAVVKIRANGTGAATVTSKBSAEATLADPEVAQASQAEISTFGG MKNIALIDFGKKSIASSLVKRGCKVTVVYQDMEEVYIKPDGIVLSNGPDEPAIO PYLGRKISISRPPTLIGICGHOLIALAGGKFMKIPFGHRANMPVINDRTKRVEMT SONSHVYDEQSIINEELTTRFHHVNDTSVEGLAHKKLIPVMSVQFHPRAHQAASEM IFDDYLNKVIIPAREIAHA"


```

(512)"
/codon_start=1
/translation=1
/protein_id="BA19701.1"
/db_xref="GI:1945093"
/translation="MIIIPNNIEIAKHMILDFEAKHMGCTPEMAISSGIFRCDELGFA
AVNESEIGICITITITIDGKDCILSIDSESRMLLEOPAPELCADRDYTLDMYRGH
ITNEHMQTEIDHLIVNGRMPDSTEAEMFVSYYKKEVIDFPLNPNVAVCYETIARA
ITNDNNNAIAFVOKRNOYCPAAVFPNVAEAKRLKEIFEIVANGLIIDELEFSKVID
"
gene
9392.10369
/gene="ydhj"
/cds
9392.10369
/gene="ydhj"
/codon_start=1
/translation=1
/protein_id="BA19702.1"
/db_xref="GI:1945094"
/translation="MEISDIYGOHHIDGVLIELIKSAPVORLKGITYOGASFLVNRK
MNTVREHSIGVWLKIKGITEIOIGILDEHSHTAFSHVDVDFENQADYENHENT
FOOVIVHSRIPDLOHGYHTEBELSDSRMLLEOPAPELCADRDYTLDMYRGH
ITNEHMQTEIDHLIVNGRMPDSTEAEMFVSYYKKEVIDFPLNPNVAVCYETIARA
ITNDNNNAIAFVOKRNOYCPAAVFPNVAEAKRLKEIFEIVANGLIIDELEFSKVID
"
Query Match 33.9% Score 320.4; DB 2; Length 36308;
Best Local Similarity 59.4%; Pred. No. le-77;
Matches 562; Conservative 0; Mismatches 381; Indels 3; Gaps 1;

QY 1 atgagacgtgaacgcttatcttcaagcctgttctcaagaagaatttggggcgggacc 60
DB 17823 ATCATGACGCATCATATTTTATGAGCGCTGCTTTAAAGAACTATGAGGAGGACG 17882
QY 61 gcttaagctga---tttgctataccatccgtcccaagaacagggagtgctgqct 117
DB 17883 AAGCTTCGTCAGCGCTTTTGCTGACGCAATACCTCACAACAAACAGGTAGTCTGGCC 17942
QY 118 ttggcggcgatcaaaatgctcaaacgctgttcaaaaggaatgataagggttcaag 177
DB 17943 GTTCTGACATGCCCTATGCTGCTCTGTAAATAATGCCCCGCTGCAGAAAGACA 18002
QY 178 ctcaagcaatgatgggaacatcaagaacattatctggaaagcttgaagggaacgcttc 237
DB 18003 CTTCATCAATATGGAAGATGATCATCCAGACATATTCGGGTTTCCGCGATGTAAGGTGT 18062
QY 238 cctctgcttaaaaaatatagatgtctgaacgaactatctgttaagtgatccgaat 297
DB 18063 CCGCTGCTGTAAGAGCTGTCGACGCCAATATGATCTCTCCGTCGAAGTCCATCTGAT 18122
QY 298 gatgaatgacaaacacacacacacacacacacacacacacacacacacacacacac 357
DB 18123 GATGATATGCAAAACGTCGACGAAATGCGACCTTGTGTAACGGAATGCTGTGTATATC 18182
QY 358 atgattgcacaaaagatcgacgaatattatattgacacacacacacacacacacac 417
DB 18183 ATGATATGCAAAAGTGTGCGGCACTAATTTTGGACATCATCATCAAGCAAAAGGAAAG 18242
QY 418 ctaactacacatgataagcgttgaagaatggaatgaagctcttgccgctgttaagtaag 477
DB 18243 TTCAAAACAAAGATAGAAACCGGTGATGGAACGGCTCTGAGCGAATTCATAAATCAAG 18302
QY 478 cagggaggaattttctctgtgcacacacacacacacacacacacacacacacacac 537
DB 18303 CAGGAGATTTCTTTTATGTCACACGCGATCCTCTTTATGTAAGGAACCCCTT 18362
QY 538 gctttgagagcagcagaactcagacacacacacacacacacacacacacacacacac 597
DB 18363 GTCCTGTAATTCAGCAAAACCTGTGATACAACTATGCGGTATACATTAATGACCGCTGT 18422
QY 598 gatgcagaagcaagctgcgagctctctgaaagaaagacatgaaatgaaatgaaatgaa 657
DB 18423 AATGACCGCGCAAAACAACTCTCATATGAAAGCAATGGAAGTCAATACAGATA 18482
QY 658 ccgtctatctcagaagcagcagcagcagcagcagcagcagcagcagcagcagcagcagc 717

```

```

DB 18483 CCGCATATCATTAAGTGCATACACCGGAAGTAAAGAACTTGTGAACCGTGCATCAT 18542
QY 718 acattgtgaatgcgcttacttctggtgggaatggaactacagatcagaagaac 777
DB 18543 GTTATGTCATCATGATATTTCTCTGATGATACAAATGGAAGATGAGCGGAGCTGCT 18602
QY 778 ttaagcagaacaaacacacacacacacacacacacacacacacacacacacacacac 837
DB 18603 TTCTCTCATATCAAACTATTTCTGCGGAGGTGTCGACCGGATCAGACATCATATA 18662
QY 838 tctgtgagatgctatcccttcaaaaagagatcagatgctgctccatgctt 897
DB 18663 AATATGATATTCAGTATGATGATGCAATGCAAGCGCTCACACTTATTTCTGCGCATTTT 18722
QY 898 ggaacttaaacctgaagatagcagaatgctatcgtccatccatc 943
DB 18723 GGAGAAATTTCAATAGAAAGACATGTAATTCATGATATCTATC 18768

RESULT 4
LOCUS BSUB004
DEFINITION Bacillus subtilis complete genome (section 4 of 21): from 600701 to 813890.
VERSION Z59107 AL009126
KEYWORDS 259107.1 GI:2632866
SOURCE
ORGANISM
Bacillus subtilis.
Bacteria; Firmicutes; Bacillus/Clostridium group;
Bacillus/staphylococcus group; Bacillus.
1 (bases 1 to 213190)
Kunst, F., Ogasawara, N., Moszer, I., Albertini, A.M., Alloni, G.,
Azevedo, V., Bertorello, M.G., Bessieres, P., Bolotin, A., Borczyk, S.,
Boris, R., Boudier, L., Brans, A., Braun, M., Brignell, S.C.,
Bron, S., Brouillet, S., Bruschi, C.V., Calikwell, B., Capano, V.,
Carter, N.M., Choi, S.K., Codani, J.J., Comerford, I.F., Cummings, N.J.,
Daniel, R.A., Denli, F., Devine, K.M., Dusterhoft, A., Ehrlich, S.D.,
Emmerson, P.T., Entian, K.D., Erington, J.M., Fabret, C., Ferrari, E.,
Fougeron, D., Fritz, S.Y., Fujita, M., Fujita, Y., Funa, S., Gallazzi, A.,
Gallier, G., Gieseler, G., Guy, B.J., Hag, K., Haisch, J., Harwood, C.R.,
Henaut, A., Hilbert, H., Holsappel, S., Hosono, S., Hullo, M.F.,
Itaya, M., Jones, L., Joris, B., Karamata, D., Kasahara, Y.,
Klaer, Blanchard, M., Klein, C., Kobayashi, Y., Koetter, P.,
Koningstein, G., Krogh, S., Kumano, M., Kunita, K., Lapidus, A.,
Lardinois, S., Lamber, J., Lazarevic, V., Lee, S.M., Levine, A., Liu, H.,
Masuda, S., Maestri, C., Medigue, C., Medina, N., Mellado, R.P.,
Mizuno, M., Moestl, D., Nakai, S., Nodack, N., Noone, D., O'Reilly, M.,
Ogawa, K., Ogihara, A., Oudega, B., Park, S.H., Patro, V., Pohl, T.M.,
Portetlelle, D., Portwilk, S., Prescott, A.M., Prescan, E., Puig, P.,
Purcell, B., Rapoport, G., Rey, M., Reynolds, S., Rieger, M.,
Rivolta, C., Roche, E., Roche, M., Rose, M., Sadate, Y., Sato, T.,
Scanlan, E., Schleich, S., Schroeter, R., Scoffone, F., Sekiguchi, J.,
Sekowska, A., Seror, S.J., Serron, P., Shin, B.S., Soldo, B.,
Sorokin, A., Tacconi, E., Takegi, T., Takahashi, H., Takemaru, K.,
Takeuchi, M., Yamakoshi, A., Tanaka, T., Terpe, P., Tognoni, A.,
Tosato, V., Uchiyama, S., Vandenbol, M., Vannier, P., Vassarotti, A.,
Viar, A., Wambutt, R., Wedler, E., Wedler, H., Wetzinger, T.,
Winters, P., Wipat, A., Yamamoto, H., Yamae, K., Yasumoto, K.,
Yoshida, K., Yoshikawa, H.F., Zumbstein, E., Yoshikawa, H. and
Danchin, A.
The complete genome sequence of the gram-positive bacterium
Bacillus subtilis
Nature 390 (6657), 249-256 (1997)
98044033
2 (bases 1 to 213190)
Kunst, F., Ogasawara, N., Yoshikawa, H. and Danchin, A.
Direct Submission
Submitted (18-NOV-1997) I. Moszer, A. Danchin, Institut Pasteur,
Regulation de l'Expression Genetique, 28 rue du Docteur Roux, 75724
Paris Cedex 15, FRANCE. E-mail: moszer@pasteur.fr
adanchin@pasteur.fr Phone: +33 (0)1 45 66 84 41, Fax: +33 (0)1 45

```

```
FEATURES             68 89 48      Location/Qualifiers
Source                1..213190
                    /organism="Bacillus subtilis"
                    /strain="168"
                    /db_xref="taxon:1423"
gene                 33.. 740
                    /gene="ydfS"
                    /function="unknown"
                    /note="similar to hypothetical proteins"
CDS                  33.. 740
                    /codon_start=1
                    /transl_table=11
                    /protein_id="CAB12373.1"
                    /db_xref="GI:2632867"
                    /db_xref="SPTREMBL:P96697"
                    /translation="MTELEVIRIVASFGILLIAERILKQQTISQMTTFDFIAATLIG
                    AIAAGLAYNTSIKPHMAISFISFVITFILSLISKNRKRFPAGDPFTVLQNGKI
                    LESNMKRMRTYLDLYNQOLREKEIFNIEETNGOLTVLRKPPDRHXTKODIM
                    IAVNQRPLIELIMDGEIENNLKONRLTESWLEELRKRDIKVETVAVLLGNGD
                    IYDYQKDHISVPMKE"
gene                 complement(755..1186)
                    /gene="ydfT"
                    /complement(755..1186)
CDS                  complement(755..1186)
                    /gene="ydfT"
                    /function="unknown"
                    /codon_start=1
                    /transl_table=11
                    /protein_id="CAB12374.1"
                    /db_xref="GI:2632868"
                    /db_xref="SPTREMBL:P96698"
                    /translation="MDFEIRKRLTETNEYKOSDEFTNLKYKAPHEPFSRSTFVY
                    DPEQDEYKPIYEMRQYEPETLITLSKTQILLKGTTPYPPAEMEVLRRIRYGEI
                    ERILQPEAETHLDIOLINGLHISYRQVETVAFNKGL"
gene                 complement(1199..1441)
                    /gene="ydgA"
                    /complement(1199..1441)
CDS                  complement(1199..1441)
                    /gene="ydgA"
                    /function="unknown"
                    /note="similar to hypothetical proteins from B. subtilis"
                    /codon_start=1
                    /transl_table=11
                    /protein_id="CAB12375.1"
                    /db_xref="GI:2632869"
                    /db_xref="SPTREMBL:P96699"
                    /translation="MPYQINIAIKINGVLTONGNIDVGPVHNSHTANSKIYGANFSL
                    GDLSPTSSLNTNGNIDVSDQDQIGNPSAPISNQT"
gene                 complement(1455..1727)
                    /gene="ydgB"
                    /complement(1455..1727)
CDS                  complement(1455..1727)
                    /gene="ydgB"
                    /function="unknown"
                    /note="similar to hypothetical proteins from B. subtilis"
                    /codon_start=1
                    /transl_table=11
                    /protein_id="CAB12376.1"
                    /db_xref="GI:2632870"
                    /db_xref="SPTREMBL:P96700"
                    /translation="MPSYINLYLKINSISGNSITIGEAAYSPTNNQKSQGTNSS
                    FGDTSPTSEYMEFLNDPVDQTSIGNSDTSVNAPPIAPPILD"
gene                 2026..2613
                    /gene="ydgC"
                    /complement(2026..2613)
CDS                  2026..2613
                    /gene="ydgC"
                    /function="unknown"
                    /note="similar to hypothetical proteins"
                    /codon_start=1
                    /transl_table=11
                    /protein_id="CAB12377.1"
                    /db_xref="GI:2632871"
                    /db_xref="SPTREMBL:P96701"
                    /translation="MKSQVESRKRLLKAAANFSGVGFHQAQVSEIYVKKAGPTQPSF
```

```
YLYFOSKRAIPAEILTDHSHVRKLTESILLENGJNATEDVSKRVLLAVETVPOFLDD
KDLIRKIGFLNPEAKQKMKDLAIYKLNLEAQRSLSTFHSELDMEIYAECLVGMIEHL
TESFLTGIRKDPASLAEEVNLITGMLPGKNDV"
join(2610..2954,3074..3091)
gene                 2610..2954
                    /gene="ydgD"
                    /complement(2610..2954)
CDS                  2610..2954
                    /gene="ydgD"
                    /function="unknown"
                    /codon_start=1
                    /transl_table=11
                    /protein_id="CAB12378.1"
                    /db_xref="GI:2632872"
                    /db_xref="SPTREMBL:P96702"
                    /translation="MISTMKVSLAVPMLAGGITKSVRPPOVEHRHRYPIPLFLV
                    TGILBAGALAMTGMKRYAIGAGVLEVVLAGAIIAHMFRRARQSVIMATQMICL
                    IVSIMIIMGYSY"
gene                 3074..3091
                    /gene="ydgD"
                    /complement(3074..3091)
CDS                  3074..3091
                    /gene="ydgD"
                    /function="unknown"
                    /note="similar to hypothetical proteins"
                    /codon_start=1
                    /transl_table=11
                    /protein_id="CAB12379.1"
                    /db_xref="GI:2632873"
                    /db_xref="SPTREMBL:P96703"
                    /translation="MAFTLEDMTSEEFEPFGMSVQNPFAKONITSGTTEKAEKSE
                    OAYEMINGRDSSNHYFMNITNDQGERMGMLTMYADPLHQRAPFYSPFLYAFNG
                    KAGIQLAQTDLDERARELGAERLALHVAHNEFTAVYLXQKNGVAMTNIRKQKLC"
gene                 complement(3705..3722)
                    /gene="exp2"
                    /complement(3705..3722)
CDS                  complement(3705..3722)
                    /gene="exp2"
                    /complement(3705..3722)
                    /gene="exp2"
                    /complement(3750..5393)
                    /gene="exp2"
                    /complement(3750..5393)
                    /gene="exp2"
                    /complement(3750..5393)
                    /codon_start=1
                    /transl_table=11
                    /product="ATP-binding transport protein"
                    /protein_id="CAB12380.1"
                    /db_xref="GI:2632874"
                    /db_xref="SWISS-PROT:P39115"
                    /translation="MKETVLTNYSYEPVQDTVEKRYHNASVQOGDIIGITKNGAGS
                    TLHITHNDLPAAQQLIRKDIKALYDQETAAASPDQTPAEKILIEKHVPLRDPH
                    QLSGGEKAKARLAKGLSDADLDLDEFTNRLDKRSQFLIQLRHNGYIILVSHR
                    YFLDSAAIKYINSLDQTLIEFGKNTSGYMRERKRRLTQQRKEKEQKAWERLPAQNT
                    GLASWSEKAHAOSTKEGKEHYHVKAKRTDAQIKSKQRLKELEKAKAPVPEYT
                    VAFSIDTTHKGTGRFLQVANTYKAFGEFTLRKNANFTIQHEKVAIIIGPNSGKTTLL
                    NIILQOETAEGSVWVSPSANGIYLTQEVFDLDEPQTEBELFENETFFARGHYOAMRH
                    LGFTAAOWTEPIKHMMSGERYKIKLMAYILEKQVLLIDEPFNHLDLPSREOLEETLS
                    OYSGKLAVSHDRYPLEKTTNSKLVISNNGIEKQINDVPSRNRREELRLKLETERQE
                    VLGKLSFMTPNDDKGYKELDOAFNELTKRIKRLDHQDKKD"
gene                 complement(5428..5450)
                    /gene="ydgF"
                    /complement(5428..5450)
CDS                  complement(5428..5450)
                    /gene="ydgF"
                    /complement(5428..5450)
                    /gene="ydgF"
                    /complement(5678..5703)
                    /gene="ydgF"
                    /complement(5678..5703)
                    /gene="ydgF"
                    /complement(5713..7089)
                    /function="unknown"
                    /note="similar to amino acid ABC transporter (permease)"
                    /codon_start=1
                    /transl_table=11
                    /protein_id="CAB12381.1"
                    /db_xref="GI:2632875"
                    /db_xref="SWISS-PROT:P96704"
                    /translation="MTDDMTKDNITNOQTLQGLKNRHHIOLAIAGCAIGTGLFLGSGKS
                    IHFGPSILFAYMITGILCFILMRSGLILSNYSFVDFQDYDYGMAAPFTGWT
                    YWFCWISIAMADLFAVGLYTOYMLPGVOWVPGILALIIILIMLATVAKLGELEEF
                    ALKVIATIALIIVIGLVMIFFGFSSTSSVSFTLMSHGGLFPGMHGFIILSPOMYVF
```

AFVGEIVGLTAGETENPEKVIIPKAINNIPVLLFVIGALLVIMSTVPMIDINPSES
PFOVPEVAVGIVGASIIINFVULTSASACNSAVSTSMVSLAKHDINAPESMAKLT

Query Match 33.98; Score 320.4; DB 2; Length 213190;
Best Local Similarity 59.48; Pred. No. 1.2e-77;
Matches 562; Conservative 0; Mismatches 381; Indels 3; Gaps 1;

QY 1 atgacgactgaacgcgttatcttccaagcctgtttccaagaagaatttggcgcgagacc 60
DB 30821 ATCATGACGATCCATATATTTTAAAGCCCTGTCTTAAAGACTATGGGAGGAGC 30880
QY 61 gctttagctga---tttggctatcacatccgttcacacgaacgaaggagtgctgacct 117
DB 30881 AACCTTGTACCCCTTTTGGCTACGCAATACCTCACAACAAAACAGTGAGTCTGGCC 30940
QY 118 ttgacgcgcatacaaatggtcaagcgttgttcaaaacgaagtataagggttcacg 177
DB 30941 GTTTCGTGACATGCCATGGCTGCTGTGMAAAATGCGCCGCTGGCAGGAAAGACA 31000
QY 178 ctacgcaaatatgagacacacacagacatttatacgaagcttgaaggagacgcttc 237
DB 31001 CTTCATCAAGTATGAAAGATCATCCAGAGATATCGGTTTCCGATGTAAGTGTTT 31060
QY 238 cctctgcttcaaaaatatagatgctgaacgaactatcgttgaagtgatccgaat 297
DB 31061 CGCTGTGCTAAAGCTGGCTGGACGCCAATATGATCTCTCCGTGCAATCCATCTGAT 31120
QY 298 gatgaatgac 357
DB 31121 GATGATATTCACAAACATGCGCGAATAATGGCGACCTTGTAACAAAGAGTGCTATAC 31180
QY 358 atgatgtccaaaagaatgacgaactattatgacacacacacacacacacacacac 417
DB 31181 ATTGATTTTTCACAAAGATGCGCGAATAATTTTGGACATCATGACACAAAGAGAGC 31240
QY 418 ctac 477
DB 31241 TTCAACACACGATATGAAACCGGTGATGGAAGGGGCTGCTGGGGAATCAAAATCAAG 31300
QY 478 cggggggggttttctatgctgaagcggtaactgttcaatgacatggaagaatcttc 537
DB 31301 CCAAGGAGATTTCTTATATGCGCCACGCGTACACTCAAGCTTATGTAAAGGAACCTT 31360
QY 538 gcttggagacgacgac 597
DB 31361 GTCCTTGAATTCAGCAAAACTCTGATACACATATCGGTATACATATGACCCCTCT 31420
QY 598 gatgcgaagacgacgacgacgacgacgacgacgacgacgacgacgacgacgac 657
DB 31421 AATGACCAAGGCGCAAAAGAACTCTTCAATATGAAAGCAATGGAATGCAATACGATA 31480
QY 658 cggctctatccagaacgacgacacacacacacacacacacacacacacacacac 717
DB 31481 CCCCAATATGATTAATGATACACCGGAAGTAAAGAGTTGTGAACCTGAGATCATAT 31540
QY 718 acatgatgaatgacgacgacgacgacgacgacgacgacgacgacgacgacgac 777
DB 31541 GTTATGTGCAATCAATATATTTCTCAAGTACAAATGGAAGATTAAGCCGCGACTGCT 31600
QY 778 ttaagcagaac 837
DB 31601 TTTCCTTCATATCAATCAATATATTTGCTGGGAGGTCTTCAAGCGATCAAGCAATCATA 31660
QY 838 tctggtgagatgctctatccttccaacacacacacacacacacacacacacacac 897
DB 31661 AATAAATGATATGATATGATATGATATGATATGATATGATATGATATGATATGAT 31720
QY 898 ggggaatttaaacactgacacacacacacacacacacacacacacacacacacac 943
DB 31721 GGAGAAATTAATTAAGAGCAATGTGATATCAATATCTCATC 31766
ESULT 5

AX006205
LOCUS AX006205 945 bp DNA PAT 24-AUG-2000
DEFINITION Sequence 6 from Patent WO9904016.
ACCESSION AX006205
VERSION AX006205.1 GI:9929073
KEYWORDS
SOURCE
ORGANISM
Bacillus subtilis.
Bacillus subtilis.
Bacteria; Firmicutes; Bacillus/Clostridium group;
Bacillus/Staphylococcus group; Bacillus.
REFERENCE
1 (bases 1 to 945)
AUTHORS Estell D.A.
TITLE Proteases from gram-positive organisms
JOURNAL Patent: WO 9904016-A 6 28-JAN-1999;
GENECOR INTERNATIONAL B.V. (NL); ESTELL DAVID A (US)
FEATURES
location/Qualifiers
1..945
/organism="Bacillus subtilis"
/db_xref="taxon:1423"
BASE COUNT 290 a 189 c 223 g 243 t
ORIGIN

Query Match 33.98; Score 319; DB 9; Length 945;
Best Local Similarity 59.48; Pred. No. 1.7e-77;
Matches 560; Conservative 0; Mismatches 380; Indels 3; Gaps 1;

QY 4 agactgaacgcgttatcttccaagcctgtttccaagaagaatttggcgcgagaccgct 63
DB 1 ATGAGCGATCCATATATTTTAAAGCCCTGTCTTAAAGACTATGGGAGGAGCAG 60
QY 64 ttagctga---tttggctatcacatccgttcacacgaacgaaggagtgctgacctt 120
DB 61 CTGCTGAGCGCTTTTGGCTACGCAATACCTCACAACAAAACAGTGAGTCTGGCCGCTT 120
QY 121 ggcgcgac 180
DB 121 TCTGACATGCGCCATGCTGCTGTCTGTAAATAATGCGCCGCTGGCAGAAACACTT 180
QY 181 agcaatattggaac 240
DB 181 GATCAAGATGGAAGATCATCCAGATATTCGGGTTCCGATGTAAGGTTCCG 240
QY 241 ctgcttacaac 300
DB 241 CTGCTGTAAGCTGCTGACGCCAATATGATCTCTCGTCAAGTCAATCTGATGAT 300
QY 301 gaatatgcaac 360
DB 301 GATTAATGCAAAACTGACGAAATAATGGGACCTTGTAACAGGATCTGTATATCAT 360
QY 361 gatgac 420
DB 361 GATTCGAATGAGACGCCGACATATTTTGGACATCATGCAACACAAAGAAAGATTCT 420
QY 421 actac 480
DB 421 AAACACCAATATGAAAGCGGTGATGGAACGGCTGCTGAGGCCAATCAAAATCAAGCA 480
QY 481 ggggaatttctatgctgaacacacacacacacacacacacacacacacacacacac 540
DB 481 GGAATTTCTTATATGTCACAAAGCGTACACTTCATCTTATGTAAGGAAACCTTGTC 540
QY 541 ttggaac 600
DB 541 CTGGAATTCAGCAAAACTGATATCAACAAATATGCGGTATATGATATGACCCCTGAAT 600
QY 601 gcaagaagcgaagcgtgcgcgacgttcatctgaaaaagacacacacacacacacacac 660
DB 601 GACCAAGGCGCAAAAGAACTCTTCAATATGAAAGCAATGGAAGTCAATGAAGATACCG 660
QY 661 tctatccagaacgac 720
DB 661 tctatccagaacgac

QY	Db	661	CANITGCAATTAAGTCATACACCGCAATTAATAAGATTTGGTAACGCTGAATCATTTGTT	720
QY	721	ttgatgtaatgcgcttaacttttcggtggggaatrgaacttalcagatcagcaagctta	780	
Db	721	TATGTGCAATCAAGATTATTTCTCATGTGACAAAGAAATTAAGCGCGAGCTGCTTTT	780	
QY	781	aagcagcaaaacccattcccttctacagtgctgaattgaagggaaggccgctatgctct	840	
Db	781	CTTTCATATCAAAACCATTTTCTGTGGGAGGTCTTGAGCGGATCAGCAATCATCAATAT	840	
QY	841	ggtgagatgctatcccttcccaaaaaggagatcatatgctgctccttaacgcttga	900	
Db	841	AATGCGATTTCAGTATCAATATGCAATACAGCGCTCACACACTTTATCTGCTCGCTTGGCA	900	
QY	901	gaattcaactcgaagatagtcagatgcttgcctccatc	943	
Db	901	GAATTTCATATGAAAGAACATGATTAATTCATGATTCCTATC	943	
RESULT	6			
LOCUS	AX006203	948 bp	DNA	PAT 24-AUG-2000
DEFINITION	Sequence 4 from Patent WO9904016.			
ACCESSION	AX006203			
VERSION	AX006203.1	GI:9929072		
KEYWORDS				
SOURCE	Bacillus subtilis.			
ORGANISM	Bacillus subtilis.			
REFERENCE	Bacteria; Firmicutes; Bacillus/Clostridium group;			
AUTHORS	Bacillus/Staphylococcus group; Bacillus.			
TITLE	1 (bases 1 to 948)			
JOURNAL	Estell, D.A.			
FEATURES	Proteases from gram-positive organisms			
	Patent: WO 9904016-A 4 28-JAN-1999;			
	GENECOR INTERNATIONAL B V (NL); ESTELL DAVID A (US)			
	Location/Qualifiers			
	1..948			
	/organism="Bacillus subtilis"			
	/db_xref="taxon:1423			
BASE COUNT	286 a	192 c	244 g	226 t
ORIGIN				
Query Match	33.0%; Score 311.8; DB 9; Length 948;			
Best Local Similarity	58.8%; Pred. No. 1.7e-75;			
Matches	557; Conservative 0; Mismatches 387; Indels 3; Gaps			
QY	1	atgagcagctaaacccgtatttttcaagcctgttttccaagaagaatttggggcggagcc	60	
Db	1	ATGAGCGAATACACCGATTYYTCTTAACGGCTGTGTTAAAGAAAATCTGGGGGGAAC	60	
QY	61	gctttatcgtat---tttgcctataccattccgtccgtacacagcaagagggagctcgtgact	117	
Db	61	GCTTTACGAGATATGATTTGGATTCACATATTCCTTCAGAAATCAACGGGGGATCTCGGGCC	120	
QY	118	tttgcgcgcgataaatgtccaagcgtttgttccaagaacggaatgataaaggagttcaag	177	
Db	121	ATTTCGCCCTCATCAAAAGSACCGAGCACTGTTGCAAAATGCGCCCGCTTAAGGAAGAC	180	
QY	178	ctcgcgaaatagtggaacatcaagacaattatcctggcagcagcttgaaggagaccgttcc	237	
Db	181	TTTGATTCAGAGTTTGGGAAAGCGCCCTGAAAGTATGTCGGCGCGTGAAGGGGGATGGCTTT	240	
QY	238	ccctcgttctcaaaaaatataagatgctgcgcagagactatcgttccagtggtgcatccgaat	297	
Db	241	CCGGTTCGACAAACGCTGCTGATGTGAAGGAAGATACGTCATTTAAAGTTCAACCTGAT	300	
QY	298	gataaataatgcgaactacatgaagaacggttgcttgaaaaaacagatcgttgatcat	357	
Db	301	GATTACTATGCGGAGAAACGAGAGGGAGACTCGGCACAGACGGAATGCTGATCATTT	360	
QY	358	attgatgtccaaaagaatgctcggagattattatttggcccaatgcaacaacaagaagaa	417	

Db	361	ATGCACTGTAAAGGAAAGACGCAAAATCATTTAGCGGATACGGCCGCTCAAAAAACCGAA	420
Oy	418	ctactacatcatgatagagcgtgtggaagaatggatgaagctctctgcccgttltgaagtgaaag	477
Db	421	CTGTGCACATGATTCACACGCGGTACTCGGAGAGGCGCTCGCTCGCAAAATCAAAATTTAA	480
Oy	478	ccggggagattttctctatgtgtccaaagcgtgactcgttcatgtcgatltgnaaaggaaattctt	537
Db	481	CCGGGTGATTTCTATATATGTGCCGAGCGGAACGCTGCACGCACTATTGTGCAAAGGGGCGCTT	540
Oy	538	gctttgagacgacgacgaactcagacacaaactcaagattatatatgatltgacgcgaana	597
Db	541	GTTTTSAGAGCTCAGAAATTCAGATGCCACATACCGGGGTGTACGATTGTGACCGTCTT	600
Oy	598	gatgcagaagagcaagctgcgcgaagttatcttgnaaaaaagagcctltgaagtgaagtggtgc	657
Db	601	GATAGCAACGGAAGTCCGAGAGAGCTCAATTTTGCCAAAGCGGTCAATGCCGCGACGTT	660
Oy	658	ccgtctatctccagaacgagatacagtlcacaatgacaaalttgagattgtgtctacaag	717
Db	661	CCCCATGTGACCGGTATATAGATGAATCGACAGATCAAGAAAAAGGAATMACCATTA	720
Oy	718	aaattgtgaagcgtctacttctcgttggggaatgaagactatcaagatcagaacgc	777
Db	721	ACATTGTCCAAAGGGAATATTTTCGTTATTAATAGGACATCAATGCGCAAGCTGAA	780
Oy	778	ttaaagcagcaaaaccattcctctctatcaagtgtgatgtgaaggagagggccgtatgac	837
Db	781	ATGGCTTGGATGAATTCCTTCTGATTGTGCACGCTGATGAGAGGAAGCGGTTTGCTCAAG	840
Oy	838	tctgtgtgagtgtctatccttccatcaaaaaagagatcatatgtgtcgtcccttaagctt	897
Db	841	TATAGGACAAACATCAATGCTCCGCTCAAAAAAGGTGATCACTTATTTTCCGCGCTCAAAAG	900
Oy	898	ggagaatttaactcgaagatcagatcagaagatgcatgcgtcccatc	944
Db	901	CCGATTTTACGATTAAGAGAACTGTACCCCTTACGTCGTCTCATAT	947
RESULT	7		
BSU02562			
LOCUS	3962 bp	DNA	BCT 03-JUN-1995
DEFINITION	Bacillus subtilis N-acetylglucosaminidase (lyd) gene, complete cds.		
ACCESSION	U02562		
VERSION	U02562.1	GI:476091	
KEYWORDS			
SOURCE			
ORGANISM	Bacillus subtilis.		
	Bacillus subtilis		
	Bacteria; Firmicutes; Bacillus/Clostridium group;		
	Bacillus/Staphylococcus group; Bacillus.		
REFERENCE	1 (bases 1 to 3962)		
AUTHORS	Margot,P., Mauel,C. and Karamata,D.		
TITLE	The gene of the N-acetylglucosaminidase, a Bacillus subtilis 168		
JOURNAL	cell wall hydrolase not involved in vegetative cell autolysis		
REFERENCE	Mol. Microbiol. 12 (4), 535-545 (1994)		
MEDLINE	95020588		
AUTHORS	2 (bases 1 to 3962)		
TITLE	Vladimir V. Lazarevic., TITLE	Direct Submission.	
JOURNAL	Direct Submission		
	Submitted (18-OCT-1993) Vladimir V. Lazarevic, Institut de		
	Genetique et de Biologie Microbiennes, Rue Cesar-Roux 19, Lausanne,		
	1005, Switzerland		
FEATURES	Location/Qualifiers		
source	1..3962		
	/organism="Bacillus subtilis"		
	/strain="168"		
	/db_xref="taxon:1423"		
	/map="310 degrees"		
	3..8		
	/note="sigmaa"		
	30..35		
	/note="sigmaa"		
	-10_signal		
	-35_signal		
	-10_signal		

Db 817 GATAGCAACGGAATCGAGAGAGCTTCAATTTGCCAAGCGGTCAATGCCGACGGTT 876
 Oy 658 ccgcataccgaagcgcatacagccatgaacaaatggagattgcttcaacg 717
 Db 877 CCCCATGTGACGGGATATATGATGAATGACAGCATCAAGAAAGAAATATACATTA 936
 Oy 718 acattgattgagccttcttcttgcgtggtgggaatgaattatcagatcaagaac 777
 Db 937 ACATTTTCCAGAGGAAATATTTTCGCTTTATTAATGGACATATGCGAAGCTGAA 996
 Oy 778 ttaagcagcaaaacaccatccttctatcatcagtgatgaaggaggccgattgac 837
 Db 997 ATGGCTAGGATGATATCTTCTGATTTGACGCGTATGAAAGGAACGCTTTCCTCAAG 1056
 Oy 838 tctgtgagatgtctatccttcaaaagagacatcatgtctgctccttcaagctct 897
 Db 1057 TATGAGCAAAACATGCTCGCTCAAAAAGGATGATCACTTATTTTCCGCGCTCAATG 1116
 Oy 898 ggaagattaaactcgaagatatgcagaatgcatcgtctcccatc 944
 Db 1117 CCCGATTTTGCATTAAGAAAGAACTTGACCCCTATGCTGTCATAT 1163

RESULT 9
 BSUB0019/c
 LOCUS Bacillus subtilis complete genome (section 19 of 21): from 3597091
 DEFINITION 259122 AL009126
 VERSION 259122.1 Gi:2636029
 KEYWORDS
 SOURCE
 ORGANISM
 Bacillus subtilis.
 Bacteria: Firmicutes; Bacillus/Clostridium group;
 Bacillus/Staphylococcus group; Bacillus.
 1 (bases 1 to 212610)
 Kunst, F., Ogasawara, N., Moszer, I., Albertini, A.M., Alloni, G.,
 Azevedo, V., Bertero, M.G., Bessières, P., Bolotin, A., Borchert, S.,
 Borriss, R., Bouvier, L., Brans, A., Braun, M., Bridgell, S.C.,
 Bron, S., Broutillet, S., Bruschi, C.V., Caldwell, B., Capuano, V.,
 Carter, N.M., Choi, S.K., Codani, J.J., Connerton, I.F., Cummings, N.J.,
 Daniel, R.A., Denicot, F., Devine, K.M., Dusterhoft, A., Ehrlich, S.D.,
 Emerson, P.T., Entian, K.D., Errington, J., Fabret, C., Ferrari, E.,
 Foulger, D., Fritz, C., Fujita, M., Fujita, I., Fuma, S., Galizzi, A.,
 Galleron, N., Ghim, S.Y., Glaeser, P., Gottlieb, A., Golligly, E.J.,
 Grandi, G., Giuseppe, G., Guy, B.J., Haga, K., Haeck, J., Harwood, C.R.,
 Henaut, A., Hilbert, H., Holsappel, S., Hosono, S., Hullo, M.F.,
 Itaya, M., Jones, L., Joris, B., Karamata, D., Kasahara, Y.,
 Klaerr-Blanchard, M., Klein, C., Kobayashi, Y., Koester, P.,
 Koningsstein, G., Krogh, S., Kumano, M., Kurita, K., Lapidus, A.,
 Lardinois, S., Lauber, J., Lazarevic, V., Lee, S.M., Levine, A., Liu, H.,
 Masuda, S., Maue, C., Medigue, C., Medina, N., Noone, D., O'Reilly, M.,
 Mizuno, M., Moestl, D., Nakai, S., Noback, M., Noone, D., O'Reilly, M.,
 Ogawa, K., Ogilwara, A., Oudega, B., Park, S.H., Parro, V., Pohl, T.M.,
 Portetle, D., Portwoll, S., Prescott, A.M., Prescan, E., Pujic, P.,
 Rivolta, B., Rapoport, G., Rey, M., Reynolds, S., Rieger, M.,
 Rivolta, E., Rocha, E., Roche, B., Rose, M., Sadale, Y., Sato, T.,
 Scanlan, E., Schleich, S., Schroeter, R., Scoffone, F., Sekiguchi, J.,
 Sekovik, A., Serot, S.J., Setro, P., Shin, B.S., Solido, B.,
 Sorokin, A., Taccani, E., Takagi, T., Takahashi, H., Takemaru, K.,
 Takeuchi, M., Yamakoshi, A., Tanaka, T., Terpestra, P., Tognoni, A.,
 Tosato, V., Uchiyama, S., Vandenbol, M., Vannier, F., Vassarotti, A.,
 Viari, A., Wambutt, R., Wedler, E., Wedler, H., Wetzinger, T.,
 Winters, P., Wipat, A., Yamamoto, H., Yamane, K., Yasumoto, K., Yata, K.,
 Yoshida, K., Yoshikawa, H.F., Zumbstein, E., Yoshikawa, H. and
 Danchin, A.

TITLE
 The complete genome sequence of the gram-positive bacterium
 Bacillus subtilis
 Nature 390 (6657), 249-256 (1997)
 JOURNAL 98040033
 MEDLINE 2 (bases 1 to 212610)
 REFERENCE Kunst, F., Ogasawara, N., Yoshikawa, H. and Danchin, A.
 AUTHORS Direct Submission

JOURNAL Submitted (18-NOV-1997) I. Moszer, A. Danchin, Institut Pasteur,
 Regulation de l'expression Genetique, 28 rue du Docteur Roux, 75724
 Paris Cedex 15, FRANCE. E-mail: moszer@pasteur.fr.
 adanchin@pasteur.fr Phone: +33 (0)1 45 68 84 41, Fax: +33 (0)1 45
 68 89 48

FEATURES
 source
 Location/Organism
 1.212610
 /organism="Bacillus subtilis"
 /strain="168"
 /db_xref="taxon:1423"
 /complement(10..3879)
 /gene="yvnb"
 /complement(10..3879)
 /gene="yvnb"
 /function="unknown"
 /codon_start=1
 /transl_table=11
 /protein_id="CAB15521.1"
 /db_xref="GI:2636030"
 /db_xref="SPTREMBL:O34986"
 /translation="MRKYTVASTILSPILSYSGHHSKAPFYVOOELFQPHYIDL
 LENPOIDVKAEMRSNLTIEKAVRDNQVNSAFPLFYKOSNELGYKMPMDPGAVR
 KFLAQIPRLIMNSLEYITVMSNGKORVSEETKKLLEGOADLHAPLITLLEAV
 DTKNIGRADGYEETIEPTDRTDIPKDYHRTYRPREGSDLIIMPDRILITPGE
 PTVMALRKAGDELTADPNRYVOTLKEGNLAVIDETGEMANTPRAPVITNGK
 DISVAHYKRLRLSSITKYKPLANGRELINISIGKNSPGCVLOAOVPRKRLK
 LDKEKPYLEDLDRKVPRAESIELNADIRDSIVKATATYTRDDEKPKRLAKRD
 RNDNFRHIVISPELIGDQLEIYVAKGDIGNEKRTVKNIDIKQTSKAGLKLAKD
 ROTLSGTOPLKATTEGRADSTIKMLIDGKQTEPRAMEKRYVAFKTDKRLKRV
 MEGKVLKVEFDITNRTYSVLPETLRLKQKQRLITISGSKVSFSDPDKRNDVL
 VNANRLVSDGTVIINDQVSPERKELFJEDNQRNSKMOFQNPDLGTLQLEMTS
 KLESAHHIQASDQKENVSLVAVDNGSPHLEPNITGEGQKQNLILQAMQKMSRT
 EYDEADPTNKAKLVSRYKSSDLPGRKSLVATDLAGNKVYIERFKTERHPQDP
 PERKEPTYNKLELERSVADQVSTENKELFPHYRFEVTPVPSIDENDLAESYKGS
 LEGRKYTVAMNYRTNEMQPVDSFVADKDFTLKASVIAADYVRESKMNVIQDEIP
 PAKMDYVMSDQYVYAESYPILEDKOTELKDNQKQLMKYVFTGDIYDSADIR
 OKNADRSNVSIDKSGIYGVIALNHQDGHGKSGYRAFGKDFGDRDKFHYGSGYK
 NNRGIVLISNGNDYIMLVGMGITDDEIDAMNQVLRKHPDRAIALIAPHEYLVSNG
 RSPIEKIEFKETLEKPNPVVWVLGSHYASARKTDELDDGDKPRLVLMQADYOG
 GPEGGQVRLILOFQANDWVASTVSPYVYQDKVYVTDYVGNKDFEISLIDKPRK
 KPEPTVPCNNTYINELKREQVSGDPAERMDHLEPQSVYIYIIVEDSPFKGRKS
 ETKKRTKKEIYRPAQDPDFRHSNP"
 4044..4517
 /gene="yyna"
 4044..4517
 /note="yyna"
 /function="unknown"
 /note="similar to hypothetical proteins from B. subtilis"
 /codon_start=1
 /transl_table=11
 /protein_id="CAB15522.1"
 /db_xref="GI:2636031"
 /db_xref="SPTREMBL:O34692"
 /translation="MDSNIDHELFQALIQOFAFKRDKRWKQVQIPSGISQHHNDHK
 KWTUQLOHIVYSCHITSONVNSFLASLHISKAAYKAAVHALKHIIITVTKPKPK
 KRIEFTYDQSGKIALAHEOLHAKAKQKQLEFNEPSIDDLKVTAFENMIKYM"
 complement(4558..5775)
 /gene="cypx"
 /complement(4558..5775)
 /gene="cypx"
 /codon_start=1
 /transl_table=11
 /product="cytochrome P450-11k enzyme"
 /protein_id="CAB15523.1"
 /db_xref="GI:2636032"
 /db_xref="SPTREMBL:O34926"
 /translation="MSOSIKLFVSILSPOFNNPAPYFSQLREDDPVHYEESIDSYFIS
 RYHDVRYLQHPDIPTTKSLVRAEPIVPGCVLAOMGRKSHAKRIRVHSEFGADLH
 HSPILIKONAEMLAPYLERKSDLVNDPQGTATVACTMNLGIDLRVHSEKSEMSG
 VADFTISQSEPARHSLKCSQSDQYLMFVYIERKYNFSGSDISLCTSEYEGNAL
 SKDILALILVNLATEPADKTLALMAYHLNPNQEMVNDVLDKRSIVPRAIKETVRY
 KPVOLIPROLSDITVGGMEIKKQITVFCMIGANRDPPEAFQDPVFNIRHDLGIK

Db 89617 CCCGATTTTACAGATAAAGCACTGTACCTTATCGTCTCATAT 89571

RESULT 10

AB011838 15012 bp DNA BCT 24-FEB-1999
 LOCUS Bacillus halodurans C-125 genomic DNA, 6A fragment, clone ALBAC004.
 DEFINITION
 AB011838
 ACCESSION
 AB011838.1 GI:4512380
 VERSION
 KEYWORDS
 SOURCE

ORGANISM Bacillus halodurans (strain:C-125) DNA, clone_11b:lamba no.4

REFERENCE 1 (sites)

1 (sites)
 Takami,H., Nakasone,K., Ogasawara,N., Hirama,C., Nakamura,Y.,
 Masui,N., Fujii,F., Takaki,Y., Inoue,A. and Horikoshi,K.
 Sequencing of three lambda clones from the genome of alkaliphilic
 Bacillus sp. strain C-125
 Extremophiles 3 (1), 29-34 (1999)
 2 (bases 1 to 15012)
 Takami,H. and Inoue,A.
 Direct Submission
 Submitted (04-MAR-1998) to the DDBJ/EMBL/GenBank databases. Hideto
 Takami, Japan Marine Science and Technology Center, Deep-sea
 Microorganisms Research Group, 2-15 Natsushima, Yokosuka, Kanagawa
 237, Japan (E-mail:takamih@jamstec.go.jp, Tel.81-468-67-3895,
 Fax:81-468-66-6364)
 Sequence location (28-Apr-1998).
 Location/Qualifiers

COMMENT
 FEATURES
 source

1. 15012
 /organism="Bacillus halodurans"
 /strain="C-125"
 /db_xref="taxon:86665"
 /clone="ALBAC004"
 /clone_11b="lamba no.4"
 /map="6A"
 /note="alkaliphilic Bacillus"
 4. .9
 /note="SD sequence"
 21. .1421
 /gene="mmge"
 21. .1421
 /gene="mmge"
 /note="similar to B. subtilis mmge gene(74%-identity)"
 /codon_start=1
 /transl_table=1
 /protein_id="BAA75344.1"
 /db_xref="GI:4512381"
 /translation="MLEQIADYVLHHEITSEAFSTAYVLLDITGGGIALNYPCT
 KHLGPIVGTIVPRGTSTYVLDVQGFNFVGMVRLDYNLAVEMQSPSN
 LGGILAVDFISQENLSKGRPLKVKDYLEMMIKAHILQGLALENSLNVGSDHYLF
 VKIATAVTKMLGSDROEITNLSNMIIDSSIRTYROAPNTGSRKSNAGATSPSA
 VRLAMATKSGEMGYOTALSAPGCFVYLFNKKOLIVNOPDSYVMENYLFKVSFPAE
 FNAQTAAALAEIEHBEVKGRLDEITITSHESAIRIIDKRGPIYAPARDICLOYI
 TAVGILGATVAEYDEDAADPRIDILRDMKVTENKDYTKYIDPKRSINAYDV
 HKQSTSTANTEROYPLGHRREBNPKIYEKAAALGTHYAKKQLEQLEAVSYDID
 RLAMNVNVERTELEFV"
 1425. .1430
 /note="SD sequence"
 1439. .2341
 /gene="yq10"
 1439. .2341
 /gene="yq10"
 /note="similar to B. subtilis yq10 gene(68-77% identity)"
 /codon_start=1
 /transl_table=1
 /protein_id="BAA75345.1"
 /db_xref="GI:4512382"
 /translation="MAWIVQPKSQKELADRFQRLMKEEAILDIPGAHDAALVAK
 AGFSALYISGDATYASRGLPDGLIVISTEYVADRKKDLVRATNLPVLVDITDGGVAIN

RBS
 gene
 CDS

RBS

gene

CDS

gene
 CDS

RBS

RBS

CDS

RBS

gene

CDS

RBS

VARTAOEMLEANVAVOIEDOOLPKKCHLNGKOLVSKEMEORQAIRKVAPTLVY
 ARTDARANGELNGAIERANVYIEAGADIPPEALQSOSEEFRLVAVENSAPLAMTFE
 GRTPLMTAGLQNGAFQWYIVPTSLRVAARAKAYERIFOLIRDECTOGAGIEDQMOTFRE
 LYETISYDDFEALDKRIAKTVLGE"
 1856. .1863
 /gene="yq10"
 /note="SD sequence"
 2939. .3196
 /gene="ydhp"
 2939. .3196
 /gene="ydhp"
 /note="similar to B. subtilis ydhp gene(59%-identity)"
 /codon_start=1
 /transl_table=1
 /protein_id="BAA75346.1"
 /db_xref="GI:4512383"
 /translation="MKTEWAPADVPISIDKLVIHISNSNSYSHDVNRQKIOTSLNKS
 DIKLIDYVRRKALYHKKADCNKDFIGLTVRECOQV"
 3446. .3451
 /note="SD sequence"
 3458. .5105
 /gene="ydhc"
 3458. .3787
 /gene="ydhc"
 /note="similar to B. subtilis ydhc gene(69%-identity)"
 /codon_start=1
 /transl_table=1
 /protein_id="BAA75347.1"
 /db_xref="GI:4512384"
 /translation="MEADFDOLSSDOJPIPMILHSNGRSKILQALREYREGKEA
 NDLEAEKDLNVAHEHFKLYQCATGTEKTELILLMHARDHLSLTITMKDVKELL
 EIFFMKK"
 3799. .3804
 /gene="ydhc"
 /note="SD sequence"
 4222. .4226
 /gene="ydhc"
 /note="SD sequence"
 4239. .5105
 /gene="ydhc"
 /note="similar to B. subtilis ydhc gene(81%-identity)"
 /codon_start=1
 /transl_table=1
 /protein_id="BAA75348.1"
 /db_xref="GI:4512385"
 /translation="MNTAFSLSEYRRTYQVQNTIKMPGYPVPAKSPALIPAVIT
 LTVFLANVFSQLEHTNMHDYIQVQAPLVGLSSGSLPILILFIQLMFRGLG
 OVITNSVMDPIWNTLMIENLEAYTNGEVPNISRPFLEITVMSGTGMFLAVFAI
 LLEPKSQDKQVKGIGIGPGLVNEPIIFGLPIVNMELIIVPMILSPMTVTLVYFIS
 NASGLVPPGTGAVPMVIVPFIINGIMATNSIAGGLQLVNLMIVEMVPEFLKSMORF
 NLKKQDERKSA"
 5104. .5111
 /note="SD sequence"
 5113. .6474
 /gene="ydhp"
 5113. .6474
 /gene="ydhp"
 /note="similar to B. subtilis ydhp gene(80%-identity)"
 /codon_start=1
 /transl_table=1
 /protein_id="BAA75349.1"
 /db_xref="GI:4512386"
 /translation="MPCMTKDVTYTRPEPGFMGSAATATQTEGATRDGRKGNIMDH
 MTEVPSRPFQVQFQDSQFERYEKEDIQMKKEVGNHSFSSISMSRLIPDGGEVN
 PKGEVFNVNDIDELGQIEPEVNLPHFMDPEDEKQIGMANREVDVAYEAVATCE
 RFGDRVKKMFTHNEPIVVEGYLYDFPNEVDEKALIOGYGHTVLSAKAIRAYQ
 LNOGKGIGIILNIPSPYRSNHPADRAAKATADEFNRSPLDVGKSPPEELVOTLK
 BEGEMPIIEEDDELEIIRHVTVDICVNYQDPRVAKETLIPDPAPEMPRYEDHYAM
 BGRKNVVRGMEIYKGIYDILINLKEHGNIEPTISDNKGVGGEKPFDENGVIND
 DRIEFIDHLMVHRAIQEGSNVKGILMTFKDMMSMANVAKNRLMWGVCQFROENG
 IDN"
 6505. .6509
 /note="SD sequence"

MEDLINE REFERENCE AUTHORS	99184645 2 (sites) Takami,H., Nakasone,K., Ogasawara,N., Hirma,C., Nakamura,Y., Masui,N., Fujl,F., Takaki,Y., Inoue,A. and Horikoshi,K. Sequencing of three lambda clones from the genome of alkaliphilic Bacillus sp. strain C-125	gene	/note="alkaliphile"
JOURNAL MEDLINE REFERENCE AUTHORS	99184646 3 (sites) Takami,H., Takaki,Y., Nakasone,K., Hirma,C., Inoue,A. and Horikoshi,K. Sequence analysis of a 32-kb region including the major ribosomal protein gene clusters from alkaliphilic Bacillus sp. strain C-125	CDS	/gene="BH3777" 101..652 /gene="BH3777" /codon_start=1 /transl_table=11 /product="pyrazinamidase/nicotinamidase"
JOURNAL MEDLINE REFERENCE AUTHORS	99209008 4 (sites) Takami,H., Masui,N., Nakasone,K. and Horikoshi,K. Replication origin region of the chromosome of alkaliphilic Bacillus halodurans C-125	gene	/protein_id="BAB07496.1" /db_xref="GI:10176402" /translation="MSKALIVIDYTNDFIADGALTCGCEMERPOAQGLNERPAQVYTGILGP QHYVYATDIIHADVDVHHPMKLEPPNLSGSGRLYSGLOEPFSIPELPVMTD KTRRSAPAGTDLELRQERGIRELHLVGVCTDVCILHTAVDANKGFSILVHERGVAS FNEVGHKWLAGEFKSSIGADVRS" complement(846..1301) /gene="BH3778" complement(846..1301) /gene="BH3778" /note="BH3778 unknown"
JOURNAL MEDLINE REFERENCE AUTHORS	99356711 5 (sites) Takami,H., Takaki,Y., Nakasone,K., Sakiyama,T., Maeno,G., Sasaki,R., Hirma,C., Fujl,F. and Masui,N. Genetic analysis of the chromosome of alkaliphilic Bacillus halodurans C-125	CDS	/codon_start=1 /transl_table=11 /protein_id="BAB07497.1" /db_xref="GI:10176403" /translation="MRGWIAACVVALTCVLCGCEMERPOAQGLNERPAQVYTGILGP GPANYDNVYKRSIAPDAREONGSPFTATETPTIATDITIRSIIDQELHRSIMV MTASSHAWVNVFEPNPKKEORNKIMNDLREDMQROIIPROYHLSQKG" complement(1382..1996) /gene="tdk" complement(1382..1996) /gene="tdk" /EC_number="2.7.1.21" /note="BH3779" /codon_start=1 /transl_table=11 /product="thymidine kinase"
JOURNAL MEDLINE REFERENCE AUTHORS	99411980 6 (sites) Takami,H. Genome analysis of facultatively alkaliphilic Bacillus halodurans C-125	gene	/protein_id="BAB07498.1" /db_xref="GI:10176404" /translation="QMIIHKEGWIEVYCGSMFSGKSEELIRRYRATYRLKVOYFKP AIDRYSISQEVASHNGKVCALIPERKMSITLQWSDTDIVLIDVQFPEDEVITGVD VLADHGVATIGADIDQFGEPEPTTILMAWEYVTKDALICPPVCSGSPASRTRLID QKPSYHDPITILVGAALAYEPRCNHCHQVYVNPKSLPEVASGE" complement(2159..2404) /gene="rpm2" complement(2159..2404) /gene="rpm2" /note="rpm2" /codon_start=1 /transl_table=11 /product="50S ribosomal protein L31"
JOURNAL MEDLINE REFERENCE AUTHORS	20263314 7 (sites) Takami,H. and Horikoshi,K. Analysis of the genome of an alkaliphilic Bacillus strain from an Industrial point of view	CDS	/db_xref="GI:10176405" /translation="MKRGCIHPDYKVKVYFMDSTGFEKLSGSKTSNTEIEMEDGNTYP LIVKVISDSHPFTGOKLADAGKVRDKKTKNLK" complement(2516..3787) /gene="rhc" complement(2516..3787) /gene="rhc" /note="BH3781" /codon_start=1 /transl_table=11 /product="transcriptional terminator"
JOURNAL MEDLINE REFERENCE AUTHORS	20426005 11 (bases 1 to 305153) Takami,H. and Takaki,Y. Direct Submission Submitted (22-MAR-2000) Hideto Takami, Japan Marine Science and Technology Center, Deep-sea Microorganisms Research Group, 2-15 Natsushima, Yokosuka, Kanagawa 237-0061, Japan (E-mail:takami@jamstec.go.jp) URL: http://www.jamstec.go.jp/jamstec-e/bio/DEEPSAR/Research.html, Tel.81-468-67-3853, Fax:81-468-66-6364)	gene	/db_xref="GI:10176406" /translation="MGVHSELNMRKURELYELARKYKISYSLKRLKELFAILKGO AEEGLMFMSEVLEITOSGEGPLRPINWPSSEDIYISASQIRREDLFCNGVSGKY RPPENRERHLLHVEVKNEDDPTSKERHFPALPLPERKLNLETPGHISRII DMISPVGGQGLIVAPKAGKTSLSKELANSIVHNHPQVILLIMDSITRLARVNDVI PSGRTLSGGIDPAAFRRKRFEGAANNIEGSLITLVALVEGSRMDVITYEFK GTGMWELHLDRKLSERRIPALDIRRSTKRELLAPKSHLEKLMVARTNMDSPEFV DHFIRYKERTKTNDPEEAMKDMTNGTRKK" 3913..4257
FEATURES Source	1..305153 Location/Qualifiers /db_xref="taxon:86665"	gene	

/codon_start=-1
/transl_table=-1
/product-"conserved hypothetical protein"
/protein_id-"AAK04859.1"
/db_xref-"GI:12723679"
/translacion-"MNTITDPLFGVLYLVFMIGVRINENFRKRWNPPLFANVYI
ILISLAHITPKNTIGGSLINDLIGPSTVALGIPYKTFHLMKHAISVSIATA
LVNCTIFALCAKFFSUSKLASISIFPKSVTAMAGISDMHGVEDQITVVVAVGIL
TSVIGGLKLFRTIDPVAGIALGGTGAHVGTALIELGKTQGAALISIGVIMY
VIAPIIAKIIILY"
complement(785. .1189)
/gene-"yhnd"
/note-"L172519"
complement(785. .1189)
/gene-"yhnd"
/note="EVIDENCE BY HOMOLOG Y BIO15.02 HYPOTHETICAL,
conserved. 50% identical to PH1801 of Pyrococcus
horikoshii."
/codon_start=-1
/transl_table=-1
/product-"conserved hypothetical protein"
/protein_id-"AAK04860.1"
/db_xref-"GI:12723680"
/translacion-"MKYIDLITFFPSFIGNVSNFRLPVPGSLGMLLELALQF
KLEFRHVDAGSFLINNMTIFLPAGVGIMAMNLSHFMAQILLIVVAALIMLI
LGKLVEMIKVYKFGDYVALKPTTIVTSIEDQ"
1462. .1935
/gene-"yhne"
/note-"L173151"
1462. .1935
/gene-"yhne"
/note="EVIDENCE BY HOMOLOG Y BIO15.02 HYPOTHETICAL,
conserved. 36% identical to ylx5 of B. subtilis."
/codon_start=-1
/transl_table=-1
/product-"conserved hypothetical protein"
/protein_id-"AAK04861.1"
/db_xref-"GI:12723681"
/translacion-"MNDYKIVYEDFIRPHIPQFELFAVEMKEKFGDMKSLIVDREG
GIEIDETAEISEIISPLDITSPDPTTEILEVASFARPLKRAEHFAGAVEYI
FVKIKQINNEKEITGDSVSEDETIVVDLDTKRNKTEIPLSAVAKQIVWK"
2100. .3248
/gene-"nusa"
/note-"L173813"
2100. .3248
/gene-"nusa"
/note="EVIDENCE BY HOMOLOG Y BIO11.02 RNA synthesis,
modification, and DNA transcription. COULD PARTICIPATES IN
BOTH THE TERMINATION AND ANTITERMINATION OF TRANSCRIPTION"
/codon_start=-1
/transl_table=-1
/product-"transcription termination protein NUSA"
/protein_id-"AAK04862.1"
/db_xref-"GI:12723682"
/translacion-"MSKEMLANAFAMLEBOKGKISELYVIAIECATITAYKROYGOAO
KVIFDEKGNFNYSITREYVDEVPDSRLEISIEDNHLNPHETIGDKIMPEEKRF
ARITAGAKOVIMOKMKEERTIIVNEYSKDELIIGTVEKVDARAVNVGRVDAL
LTKRDOQIPQENFVGDGRVYVADVVLTPKGTKEVLSRAPMLKMEFEKEIPEVFDG
TVEIKSTARAGGRAKIVQSHDENVAIGTWTGSRIOGIIIRLAEKEDJIEWS
EDKATIVANALKPARIEOVLITADGSSLAIVARDOQLSLAIGRQGVNRVIAAHVTSKI
DIRKADEFNLDDYEFGEDEEIPTEVENETIVEVEVTPRESESE"
3414. .3740
/gene-"yhng"
/note-"L175136"
3414. .3740
/gene-"yhng"
/note="EVIDENCE BY HOMOLOG Y BIO15.01 HYPOTHETICAL, General"
/codon_start=-1
/transl_table=-1
/product-"HYPOTHETICAL PROTEIN"
/protein_id-"AAK04863.1"
/db_xref-"GI:12723683"
/translacion-"MKOKTPMKRSLVSNQPPKDKLRIAYNKEGEISIDPSGRKAG

RGAVTIALNKEAREAKKRVFDRAPOTKLNADFFYDELITYEHLVARELESATYASD
LAPDYD"
3743. .4048
/gene-"ylixo"
/note-"L175450"
3743. .4048
/gene-"ylixo"
/note="EVIDENCE BY HOMOLOG Y BIO12.05 Ribosomal proteins:
synthesis and modification. 45% identical to ylixo of
B. subtilis, supposed to be a ribosomal protein. Located
close to the genes implicated in protein biosynthesis."
/codon_start=-1
/transl_table=-1
/product-"probable ribosomal protein"
/protein_id-"AAK04864.1"
/db_xref-"GI:12723684"
/translacion-"MDKQOISNLGIAKRAKTIISGEELVYKSIQSSKATLVFLAND
AANLSKRINDKSNYEIAVSTLFEIELSHAIQGNRYIALYVDGFPKKMESLMNN"
4069. .6903
/gene-"l0372"
/note-"L0372"
4069. .6903
/gene-"l0372"
/note="EVIDENCE BY HOMOLOG Y BIO12.05 Ribosomal proteins:
synthesis and modification. 45% identical to ylixo of
B. subtilis, supposed to be a ribosomal protein. Located
close to the genes implicated in protein biosynthesis."
/codon_start=-1
/transl_table=-1
/product-"translacion initiation factor IF-2"
/protein_id-"AAK04865.1"
/db_xref-"GI:12723685"
/translacion-"MSDKRRINOIAKETGLTNAELVSAAOULGFEVKSSSHSSVYAEQA
EKIIOASKTQDTQTAKEPVKESQPKAESAKKNEDEHPTFPFGKAVVEDPALAR
IKAKEEKAKAYEVAESTEPVYTERPKASEPVKAEPEVKAESPEKVEYEDKQNTA
TSAEYKPEVNAVKKEPEVTEERKESLQKPKIOTIKVIRAKIDKEQAAAPERKKT
DKARNDRNNRNDNRPNONGGNGHDXDKNSGOGOGOKRQKRFAPSSAPATDS
FTPATSGKTSRRDRDRKSDNNNDNTKDGNRKGPLRVANDNRYNRNANNNMOKG
RGVONNOSSVYATQRPHELPESLEYEVGNVODIKSIRREPAETIKKLFMMGT
VNONOSIDPEPTITLMDKGYTPVKVEPEKSDIRITPEQYKLEPMNVBRPAVNT
MGHNDKRTTLIRPRESRTTEBEGAGTIOHGAOITANCKKITFILTPEGHEAPNTS
NARASVTDITILVYVADDSVMPQITELINHSKAGVPIVAINKIDKPGAFORVYIO
ELTBGVFPYAMPPENGSEVEISAKFNONDELIDTLYLVAEVDLADFTVPAVIG
LVEARLDQGGAGATLILVOQGTILHODPIVAGNTIGRVRHTNDGNIRKIKGEPSTPI
ELTGSDVPQAGHFAVFEDEKARAGERRKROQLIKONTNRNWNLSIDNLEBULKEG
QTSNVNIIIVADYQSAEALIASLQKIEVGKVIDIVSHAVGAISDSISLAASNAI
IGENNAVPTGLAREOAOEVRDIRLSIITIKYIEVEFAMRGMLOPKEKEETIGEAIV
RETFNVSKVTGIAGFVIRGKVARDAVSIVIEGVVINDGAIYSLAKHRDKVKEGMA
QEGILWEDENDEVIDDTEFYKMEVIERK"
7173. .7532
/gene-"rbeA"
/note-"L0375"
7173. .7532
/gene-"rbeA"
/note="EVIDENCE BY HOMOLOG Y BIO06.12 Sugars"
/codon_start=-1
/transl_table=-1
/product-"ribosome-binding factor A"
/protein_id-"AAK04866.1"
/db_xref-"GI:12723686"
/translacion-"MGNSEFSDVAVEIOREINDLRKNKVRPDVQVNTIDVOLTGD
LSQATVYSLSLASDNEKAVALKKAATGLKSELAKEMLTKIPLDLTFANDESVEY
GSRIDELRGLINKSEY"
7711. .8658
/gene-"pml"
/note-"L179409"
7711. .8658
/gene-"pml"
/note="EVIDENCE BY HOMOLOG Y BIO06.12 Sugars"
/codon_start=-1
/transl_table=-1
/product-"mannose-6-phosphate isomerase"
/protein_id-"AAK04867.1"
/db_xref-"GI:12723687"
/translacion-"MKEPLFNSVLOEKITGWDHKEKFGYDIPSDKVEGYAISAHPH
GVSTIANGEKKGKIDQLVASHRELFGDSKRVFPILTKIILDANWMLSVQVHPDDEG
QKHEGELGTECWYIISABGAEIIVGHNAKREBELAEIMISGDMDHLLRKYKVTGD
FFHVPSGTMAIAGIIVILETQSSDPTVRYVDFPRKNDQGNLRELHAIQOISIDVNLNP

```

GDKVPENQFKTEKPADAKITTLVKNPFDVYKMOIHGDHEFTKVADYTLVSLDGGQK
LRTDGENYEVKEGAFHILPSNEKMNLSGGLEIIASNA"
gene
/ gene="9517"
/ note="1180805"
/ gene="9517"
/ gene="img"
/ note="EVIDENCE BY HOMOLOG Y BIOD09.08 Marr-family
regulators. belongs to the Marr family"
/ codon_start=1
/ transl_table=11
/ product="transcriptional regulator"
/ protein_id="AAK04868.1"
/ db_xref="GI:12723688"
/ translation="MKTDPEKVENWLIQFHDPMLEPFLKDSAYSQVTVKELOILT
LVHTLGTGRATDIANKOKLALSTITITLNRLEDGYLERKSTADRVTHILLISKGD
ELCATRPEFPHTITDNLFDVYGTENTLAEIATLHDSLENNK"
9517..10494
/ gene="fabH"
gene
Query Match 29.38; Score 277.2; DB 1; Length 10607;
Best Local Similarity 57.38; Pred. No. 7.9e-66;
Matches 542; Conservative 0; Mismatches 398; Indels 6; Gaps 2;
OY 1 atgcagactgaacccgttatttttcaagccctgtttcaagaagaatttgggcgagacc 60
DB 7708 ATTATGGAAGACCATGTTTGAACCTCAGTTTTCAGAAAAAATCTGGCGCGGAC 7767
OY 61 gcttaagctgatttgggtatcacatcctgcacacagagggaggtgctgagcttt 120
DB 7768 CATTTGAAGAGCTTTGCTATGATTTGCCATCAGACAAAGTTGGTAATTTGGGCTATT 7827
OY 121 ggcgcgcctcaaaatggtcaaaagcgttgttcaaaacggagatgataagggttcacgtc 180
DB 7828 TCTGCTCATCCACATGCGTGTCTAACATTTGCTAATGCGAATTTAAAGGTCAAAAACTT 7887
OY 181 agcgaattatggaacacacagacattatcgcagcacttgaaggagccgtttccct 240
DB 7888 GACCAATTATACGCAAGTCACCGGAAATGTTGTGATAGTAAAGAAACAAAGTTTTCCTC 7947
OY 241 ctgcttcaaaaatcatctagatgctgacaggaacttctctgctcaggtgctcagatgat 300
DB 7948 TTACTAATCTAAATTTAGATGCAATGACTGGCTTCTGTGCAAGTTCACATGATGAT 8007
OY 301 gaatatgccaacatcatgaaacagcgtgagcttgaagaagaagaagtcgtgacattat 360
DB 8008 GAATATGACAAACATGAA--GGTGAACCTTGAAAAACTGAATGTTGTAATATT 8064
OY 361 gatgcaaaaagatgcccagattattatgcccacaaatgcaacaagaagaacta 420
DB 8065 TCACTGTAACACGCGCTGAGATTATCTATGACATTAAGCTAAATCAGTGAAGAAATTA 8124
OY 421 actacacatagatagcgctgagaaatgagatcttgcgcgctgtaagtaagcg 480
DB 8125 GCAGAAATGATTAATCTGTGATGGGATCAATTTGTAAGGTAAAGTGAACA 8184
OY 481 ggggatttttctctatgccaagcgtaacgttcaatgagatggaagaagatcttgc 540
DB 8185 GGGGATTTCTTCATGTTCCGTGAGAAATGACGATCGATCGTCTGGAATTTGTTT 8244
OY 541 ttggagagcgaggaactcaagacacacactcaagatitaaattatgacgaaagat 600
DB 8245 CTGGAACACACATCTCTGATACAACTTACCGTGTATATATTTCGACCCGTAAAT 8304
OY 601 gcagaaggaagctgcgagatctcatcgaaaaagagcatggaagtatagagtcg 660
DB 8305 GACCAAGGAATTTACGGAATTTACATTAATCAATTTGATTTATTAATTTTCG 8364
OY 661 tctattcagaac--ggcatacagttcaccatgacaatitgaggaattgtctacaag 717
DB 8365 GGCGCAAAAGTCCTGAATTTAAATTTAAATTTGCTGATGCAAAAAATTTACA 8424
OY 718 acattgatgaaatgcgttaacttgcgttgaggaaatggaactatcagatcagcaagc 777

```

```

DB 8425 ACTCTTGGAATCAAAATTTCTTGATGATTATTAATGCAAAATCATGATGACATGAA 8484
OY 778 ttaagcgcaaaaacccctctctctatcatcagtgatgataagggagcgctatgac 837
DB 8485 TTTACCAAGTTTGCTGATACATTTAGTACTTGTATGATGATGATGATGATGATGATGAT 8544
OY 838 tctgtagatgctctatccttcaaaaaagagatcatatgttgcctcctagctt 897
DB 8545 ATTGATGGAATTAATTAATCAATTTGAAAAAGAGCTCATTTTCATTTACCAACACATTT 8604
OY 898 ggaagaattaaactcgaagatgacagatgatactgctccatc 943
DB 8605 GAAATGGAATTTGCTGCAATTTGAAATTTATTTGCCAGCAATC 8650
RESULT 13
STRPM1 1379 bp DNA BCT 04-FEB-1999
LOCUS S.mutans pmi gene for mannosephosphate isomerase (complete cds) and
DEFINITION scrk gene for fructokinase (partial cds).
ACCESSION D16594.1 GI:451214
VERSION D16594.1
KEYWORDS fructokinase; mannose-6-phosphate isomerase; mannosephosphate
SOURCE Streptococcus mutans (strain:GS-5) DNA.
ORGANISM Bacteria; Firmicutes; Bacillus/Clostridium group; Streptococcaceae;
Streptococcus.
1 (bases 1 to 1379)
REFERENCE
AUTHORS Sato Y.
TITLE Direct Submission
JOURNAL Submitted (05-JUL-1993) to the DDBJ/EMBL/GenBank databases. Yutaka
Sato, Tokyo Dental College, Dept. of Biochemistry, 2-2 Masago
1-chome, Mihama-Ku, Chiba, Chiba 261, Japan (Tel.:043-279-2222,
Fax:043-279-2052)
2 (bases 1 to 1379)
AUTHORS Sato Y., Yamamoto Y., Kizaki H. and Kuramitsu H.K.
TITLE Isolation and sequence analysis of the pmi gene encoding
phosphomannose isomerase of Streptococcus mutans
JOURNAL FEMS Microbiol. Lett. 114 (1), 61-66 (1993)
MEDLINE 94123951
REMARK Erratum: [[published erratum appears in FEMS Microbiol Lett 1994 Aug
1:121(1):130]]
3 (sites)
REFERENCE
AUTHORS Sato Y., Yamamoto Y., Kizaki H. and Kuramitsu H.K.
TITLE Isolation, characterization and sequence analysis of the scrk gene
encoding fructokinase of Streptococcus mutans
JOURNAL J. Gen. Microbiol. 139 (Pt 5), 921-927 (1993)
MEDLINE 93329360
COMMENT Submitted (05-JUL-1993) to DDBJ by:
Yutaka Sato
Department of Biochemistry
Tokyo Dental College
1-2-2 Masago, Mihama-ku
Chiba, Chiba 261
Japan
Phone: 043-279-2222 x2703
Fax: 043-279-2052.
FEATURES
source location/Qualifiers
1..1379
/organism="Streptococcus mutans"
/strain="GS-5"
/db_xref="taxon:1309"
1..72
/ gene="scrk"
<1..72
/ gene="scrk"
/EC_number="2.7.1.4"
/note="3' end of scrk"
/codon_start=1
/transl_table=11
/product="Fructokinase"

```

```

/protein_id="BA04020.1"
/db_xref="GI:451215"
/translation="PAAVENSATLGLAKKIAAR"
RBS
230..234
/feature="ribosome binding site"
241..1191
/feature="pm1"
241..1191
/feature="pm1"
/EC_number="5.3.1.8"
/codon_start=1
/transl_table=1
/product="Mannosephosphate isomerase"
/protein_id="BA04021.1"
/db_xref="GI:451216"
/translation="MEGPIPILOSMHKKIMGNLRKEFGYDIPSEETGEVNAISAP
NGSVYKGVYKGVDELTAHRELFGENKSSVEPIITLIANMISVOYAPDNY
ALEHGEIGTECEYVISADEGAETIYGHAKSKEELROMLAKRDLILITKIPYAG
DEPYRSTPMHISRGIMILETOOSSEPTVRYVDPKRDQGRKRLHEQSIDVYLI
CKRANATPMHISDGLERTVLSSEPTVRYVDPKRDQGRKRLHEQSIDVYLI
TGEQYALNRKGDHLIPNTIKSMQFDLEILASHNEC"
stem_loop
1248..1289
/function="putative transcription terminator"
-35_signal
1309..1314
-10_signal
1332..1337
BASE COUNT 438 a 215 c 300 g 426 t
ORIGIN
Query Match 28.1%; Score 265.8; DB 3; Length 1379;
Best Local Similarity 57.4%; Pred. No. 9.2e-63;
Matches 539; Conservative 0; Mismatches 392; Indels 8; Gaps 3;
6 gactgaccggtatttttcaagccggtttcaagaagaatttggcgagaccgctt 65
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
244 GAAGGACCTTTGTTTAAATCACAATGCAATAAAATCTGGGCGGCAATCGGCTC 303
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
66 ag-ctgatttggcattccattccgcaacaacagggaggtcgtggtttggc 123
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
304 AGAAAAGATTGGTTTATCATCTCTGCTGCAACGAGGAGATATTGGCGATTCT 363
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
124 ggcgcatcaaatggtcacaagcgtttcaaaacggaatgtataagggttcaagctcagc 183
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
364 GCCCATCTTAATGAGAGTCTGTGTTTAAATGAGTGTAAAGGTCCTTCATGAT 423
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
184 gaattatggaaacatcacagacatttattcgagacgcttgaaggagaccgttccctctg 243
||||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
424 GAATTATATGACAGACATGAGACATGTCGSCATATACCAAGATCTGTTTCCGCTT 483
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
244 cttaacaaataattagatcgtaccaggaactatctgtcagtgatccgacgaatgataa 303
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
484 TTAACTAAATCTTAGATGCCAAGATGGGCTTAGTTCAAGTTCAATCCGATATGCT 543
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
304 taatgcaacatcatgaaaggtgaagcttggaaacagaatcgtgtacattatgat 363
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
544 TAGGCTTTGGAACATGAA--GATGAGTAGAAGACAGATGATGATGATGATGATGATGAT 600
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
364 tgcacaaagatgacgaagatttatttgaagccaatgcaacacaaaggaagaactact 423
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
601 GCTGATGAAGAGGCTGCAAAATTAATCTATGACACGAAAGCTAAATGAAGATTCGCT 660
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
424 accatataagacgttggagaatgagatcgtctgcccgtgtataaaggtataaagccggg 483
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
661 CAATATGATGCGCGGGGATGGGATCACTTTTGACAAAGATTCCTGTTAAAGCTGGG 720
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
484 gatttttctatgctcgaagcgttactgtcattcgtatgtaaaaggaatcttgcgttg 543
||||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
721 GATTCTTCTATGCTTCTAGTGGACATGATGATGATGATGATGATGATGATGATGATGAT 780
||||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
544 gagaagcgaactcagacacacacacacacacacacacacacacacacacacacacacacac 603
||||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
781 GAGACACACACACACACACACACACACACACACACACACACACACACACACACACACACACAC 840
||||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||

```

```

604 gaagcgaacgctgcgcgagcttcacatcagaagaagacatctgaatgataaggtccgcgtc 663
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
841 CAGGACGCAAAACGAGCTTTCACATGCAATGCAATGATGATGATGATGATGATGATGATGAT 897
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
664 attcagaacgcatcacagcttcaccatgaaacaaattgaggattgtcttaacaacagatctg 723
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
898 AAGCCGCAAAATGCGACCCGACGCTTGGTTATCTTACAAAGCGCTTGCAAAACGATCTTG 957
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
724 attgaatgcgttacttctcgttggggaatgagactatcagatcagaacgaactaaag 783
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
958 GTTCCAGACCCCTTCTTACTGCTATTAATGCGAGATGATGATGATGATGATGATGATGATGAT 1017
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
784 gagaacaaacacatcctcttactcagctgtaaggaagggagccgtatgactctgtg 843
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
1018 CAGACAGCTCCTTACTTCTTGGTGAATGCTTCTTCTGACAGAGTGTATTAAGCTTGGT 1077
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
844 ggtatgctatcctcttcaaaaagagatcatatgttgcctcctcaggtcttgagaaga 903
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
1078 TTAGAACATATGCTTAAAGAAAGAGATCATCTTATCTTCTTAACTATTAATCATCT 1137
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
904 tttaactcgaagataltgcagaatgtatcgtcccat 942
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
1138 TCGCAATTTGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 1176
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
RESULT 14
AC027136 49897 bp DNA HTG 30-JAN-2001
LOCUS Staphylococcus aureus clone sabac-106, WORKING DRAFT SEQUENCE, 1
DEFINITION
ordered pieces.
AC027136
AC027136.5 GI:8567822
VERSION HTG: HTGS PHASE3: HTGS_DRAFT.
KEYWORDS Staphylococcus aureus.
SOURCE Staphylococcus aureus.
ORGANISM Bacteria; Firmicutes; Bacillus/Clostridium group;
Bacillus/Staphylococcus group; Staphylococcus.
1 (bases 1 to 49897)
Loh, P., Qi, S., Ray, L., Ford, B., Worrell, V., Iandolo, J. and Roe, B.A.
Staphylococcus aureus BAC Clone sabac-106
Unpublished
2 (bases 1 to 49897)
Loh, P., Qi, S., Ray, L., Ford, B., Iandolo, J. and Roe, B.A.
Direct Submission
Submitted (28-MAR-2000) Department Of Chemistry And Biochemistry,
The University Of Oklahoma, 620 Parrington Oval, Room 208, Norman,
OK 73019, USA
On Jun 15, 2000 this sequence version replaced gi:8225121.
* NOTE: This is a 'working draft' sequence. It currently
* consists of 1 contigs. Gaps between the contigs
* are represented as runs of 'N'. The order of the pieces
* is believed to be correct as given, however the sizes
* of the gaps between them are based on estimates that have
* provided by the submitter.
* This sequence will be replaced
* by the finished sequence as soon as it is available and
* the accession number will be preserved.
1 49897: contig of 49897 bp in length.
FEATURES
location/Qualifiers
source
1..49897
/organism="Staphylococcus aureus"
/db_xref="taxon:1280"
/clone="sabac-106"
/clone_lib="OU staph library"
BASE COUNT 15753 a 9237 c 7267 g 17640 t
ORIGIN
Query Match 26.7%; Score 252.2; DB 71; Length 49897;
Best Local Similarity 35.1%; Pred. No. 7.6e-59;
Matches 516; Conservative 0; Mismatches 418; Indels 3; Gaps 1;
7 actgacacgttatttttcaagcgttcttcaagaagaatttggcgagaccgcttca 66

```



```

Db      864  CAAATCATCGAATATTTCAATATGAAATTTATGATATCATAGAAAACTGATAATAGTGAG 805
OY      613  ctgcgcgagctcaccctcgaagaaagagcattgaaagtacagagcccgctcattccagaa 672
Db      804  GAACGTGAATTAATATATAGATAAGCATTAAGATGTTATTAATTAATTAATCAATGAACCTACT 745
OY      673  cggcatcacagttcacacatgacacaattgaaagattgcttacacagacattgaaatgc 732
Db      744  AATATCATCTCCTCAAAATGAACTGATAGAAAATCACAATTTGATACACATATTTGATAGT 685
OY      733  gcttactcttcggtcgggaaatgaaactatcagatccagcacttaagcagcaaaaa 792
Db      684  GATTTTCTTACTATAGTAAAGTGAGATATTCTGTGACTCTAAATTAATTAATGAAGCTTGA 625
OY      793  ccattccctctatcacgltgattgaaggaggccgtatgatccctcgtcgtgaaatgctc 852
Db      624  GAATTTTGTCTGTCTGTCTTTAGATGACAAAGTAACCTAATGATGATGATGATATA 565
OY      853  tacccttcaaaaagagagatcatalgttcgcttaccggtcttgagagaaatlaactc 912
Db      564  TATGATATATCTAAAGTTCAAACTTTGCTTAACCTCTGAAGATTAGATAGTGTTC 505
OY      913  gaaagatatgcagaatgatcgt 935
Db      504  GAAGGAGATTTTAACTAATCAT 482

```

Search completed: May 17, 2001, 15:38:04
 Job time: 11435 sec

OK of: US-09-462-846-2 to: GenEmbl:* out-format : pfs
Date: May 17, 2001 4:11 PM
About: Results were produced by the GenCore software, version 4.5,
Copyright (c) 1993-2000 Compugen Ltd.

Command line parameters:

-MODE=framed-p2n.model -DEV-xlp
-O/cgml2/usrtio.spool/us09462846/runat_17052001_115428_937/app-query.fasta.1.376
-DB=GenEmbl -QMT=fastap -SUFFIX=p2n.rge -GAPO=12.000
-GAPEXT=4.000 -MINMATCH=0.100 -LOOPECL=0.000 -LOOPEXT=0.000
-FGAPO=4.500 -CGAPEXT=0.050 -XGAPO=10.000 -XGAPEXT=0.500
-FGAPEXT=6.000 -FGAPEXT=7.000 -YGAPO=10.000 -YGAPEXT=0.500
-DELOP=6.000 -DELEXT=7.000 -START=1 -MATRIX=blomsum62
-TRANS=human40.cdi -LIST=45 -DOCALLIGN=200 -THR_SCORE=pcr
-THR_MAX=100 -THR_MIN=0 -ALIGN=15 -MODE=LOCAL -OUTEXT=pfs
-NORM=ext -MINLEN=0 -MAXLEN=200000000
-USER=US09462846.ecgml_1.5014 -NCPV=6 -ICPV=3 -LONGLOG -NO_XLPHY
-WAIT -THREADS=1

Search information block:

Query: US-09-462-846-2
Query length: 315
Database: GenEmbl:*
Database sequences: 1283235
Database length: 1216004940
Search time (sec): 1159.440000

score list:

seq_id	Strd Orig	ZScore	Escore	len	Documentation
gb_pati:AX006200	+ 1692.00	2627.47	3.9e-138	945	AX006200 Sequence 1 from Patent
gb_ba2:BSU00007	+ 1692.00	2574.21	3.3e-135	126750	1 AB011838 Bacillus subtilis com
gb_ba1:AB011838	+ 988.50	1505.20	1.3e-75	15012	1 AB011838 Bacillus halodurans
gb_ba2:AP001520	+ 988.50	1476.08	5.3e-74	305133	1 AP001520 Bacillus subtilis N-ace
gb_pati:AX006203	+ 978.50	1516.56	3.0e-76	948	1 AX006203 Sequence 4 from Patent
gb_ba2:BSU02562	+ 978.50	1502.52	1.8e-75	3962	1 AX02562 Bacillus subtilis gene
gb_ba2:BAC0019	+ 978.50	1452.00	2.5e-73	21610	1 D59122 Bacillus subtilis com
gb_ba2:BSU00019	+ 958.50	1449.95	1.6e-74	945	1 AX006205 Sequence 6 from Patent
gb_pati:AX006205	+ 958.50	1432.83	1.5e-72	36308	1 D88802 Bacillus subtilis DNA
gb_ba2:D88802	+ 958.50	1432.83	1.4e-71	213190	1 D59107 Bacillus subtilis com
gb_ba2:BSU00004	+ 878.50	1357.01	2.3e-67	1379	1 D16594 S.mutans pmi gene for m
gb_ba3:SPRPM1	+ 867.00	1319.37	2.8e-65	10607	1 AE006310 Lactococcus lactis s
gb_ba2:SPRPM1904	+ 839.50	1250.34	2.0e-61	147006	1 AL449926 Streptococcus pneum
gb_hg12:AC027136	+ 819.50	1230.42	2.6e-60	49897	1 AC027136 Staphylococcus aureu
gb_hg17:AC078830	+ 719.00	1075.04	1.3e-51	44449	1 AC078830 Staphylococcus aureu
gb_hg17:AC074320	+ 719.00	1072.91	1.5e-51	55389	1 AC025532 Staphylococcus aureu
gb_ba1:AC078934	+ 719.00	1072.91	1.5e-51	55390	1 AC078934 Staphylococcus aureu
gb_ba2:AF269779	+ 685.00	1100.17	4.6e-53	2815	1 AF269779 Staphylococcus epide
gb_pati:AF64844	+ 718.00	1051.57	2.3e-50	2111	1 AF64844 Sequence 1 from Patent
gb_ba2:AF269301	+ 580.50	884.02	5.0e-41	3473	1 AF269301 Staphylococcus epide
gb_ba1:AB016431	+ 478.00	723.65	4.3e-32	3754	1 AB016431 Staphylococcus aureu
gb_ba1:AE001862	+ 426.00	603.58	2.1e-25	213732	1 AE001862 Deinococcus radiod
gb_ba2:AF270310	+ 419.00	635.97	3.3e-27	2433	1 AF270310 Staphylococcus epide
gb_ba1:AE001744	+ 391.00	571.36	1.3e-23	21343	1 AE001744 Thermotoga maritima
gb_ba1:AE001955	+ 213.50	305.80	8.1e-09	14878	1 AE001955 Streptococcus sanguis
gb_ba1:AE001104	+ 208.50	290.79	5.0e-08	14878	1 AE001104 Archaeoglobus fulgid
gb_ba2:DLA03118	+ 166.00	222.31	5.2e-05	3883	1 AE003118 Lactococcus lactis J7
gb_ba1:AE001146	+ 166.00	222.31	0.0004	18604	1 AE001146 Borrelia burgdorferi
gb_ba3:SCG4	+ 163.00	245.89	1.8e-05	40744	1 AL109950 Streptomyces coelic
gb_pati:AF015751	+ 160.00	210.06	0.0017	619	1 AF015751 Lactococcus lactis cy
gb_pati:AX047314	+ 145.50	195.08	0.0119	11661	1 AX047314 Sequence 28 from Pat
gb_pati:AX047311	+ 145.50	194.03	0.0136	12766	1 AX047311 Sequence 25 from Pat
gb_ba2:ECOMANA	+ 145.00	213.32	0.0012	16914	1 M1580 E.coli manA gene encodi
gb_ba1:AE000257	+ 145.00	194.77	0.0124	10914	1 AE000257 Escherichia coli K12
gb_ba2:D90805	+ 145.00	191.40	0.0191	15465	1 D90805 E.coli genomic DNA, Kc
gb_ba2:D90805	+ 145.00	189.30	0.0250	19211	1 D90805 E.coli genomic DNA, Kc
gb_ba1:AE005385	+ 139.00	185.64	0.0401	10679	1 AE005385 Escherichia coli O15
gb_ba3:MY20B11	+ 133.00	164.45	0.6063	36330	1 Z95121 Mycobacterium tubercu

gb_p13:SCYJ1036W	+ 116.50	163.17	0.7148	2912	1 Z49311 S.cerevisiae chromos
gb_ba3:SPRPM1PHI	+ 115.00	166.33	0.4767	1650	1 X57117 S.typhimurium pmi ge
gb_ba1:AE0004115	+ 114.50	146.26	6.25	12116	1 AE0004115 Vibrio cholerae C
gb_p13:SCORF512	+ 112.00	156.88	1.60	2703	1 Z48229 S.cerevisiae tRNA-Va
gb_pati:AX065227	+ 110.00	160.81	0.9672	1305	1 AX065227 Sequence 353 from

seq_name: gb_pati:AX006200

seq_documentation block:

LOCUS AX006200 945 bp DNA
DEFINITION Sequence 1 from Patent WO9904016.
ACCESSION AX006200
VERSION AX006200.1 GI:9929071
KEYWORDS
SOURCE
ORGANISM
Bacillus subtilis.
Bacteria; Firmicutes; Bacillus/Clostridium group;
Bacillus/Staphylococcus group; Bacillus.
1 (bases 1 to 945)

REFERENCE

1 Estell,D.A.
Proteases from gram-positive organisms
JOURNAL Proteases from gram-positive organisms
PATENT: NO 9904016-A 1 28 -JAN-1999
GENECOR INTERNATIONAL B V (NL); ESTELL DAVID A (US)

FEATURES

source
1..945
/organism="Bacillus subtilis"
/db_xref="taxon:1423"
BASE COUNT 288 a 180 c 232 g 245 t
ORIGIN

alignment_scores:

Quality: 1692.00 Length: 315
Ratio: 5.371 Gaps: 0
Percent Similarity: 100.000 Percent Identity: 100.000

alignment block:

US-09-462-846-2 x AX006200 ..
Align seg 1/1 to: AX006200 from: 1 to: 945

1	MetThrGluProLeuPheLeuPheProValPheLeuGluArgIleTr	17
1	ATGACGACGTGACCGTATTTTCAAGCCGTTTCAAGAAAGAAATTG	50
17	PGIYGLYThraLeuAlaAspPheGlyTrpIleProSerGlnArgT	34
1	GGGGGGGACCGCTTACCTGATTTGGCTACCATCGTCACACGAA	100
34	hrgIyGluCysTrpAlaPheAlaIahIsgIAsnGlyGlnSerVal	50
1	CAGGGGAGTGTGGCTTTTGGCCCGCATCAAAATGTCAAAGCGTTGT	150
51	GlnAsnGlyMetTrpLeuGlyPheTrpLeuSerGlnLeuTrpGluH	67
1	CAAAACGAAATGTAAGGGGTTCACCGTCAGCAATATGGAACATCA	200
151	SArgHisLeuPheGlyGlnLeuGlyAspArgPheProLeuThrL	84
1	CAGCATTTATTCGACACCTTGAGGGGACCTTCCCTGCTGCTTACA	250
201	CAGCATTTATTCGACACCTTGAGGGGACCTTCCCTGCTGCTTACA	250
84	ysIleLeuAspAlaAspGlnAspLeuSerValGlnValHisProAsn	100
251	AAATATTAAGATGCGACGACCTTATCTGTTCAAGTGCATCGAATGT	300
101	GluTrpAlaAsnIleHisGlnAsnGlyGluLeuGlyLysTrpGluCys	117
1	GAATATGCAACATACATGAAACGAGTGCATGAAACGAAATGCTG	350
301	PTyrlIleLeuAspCysGlnLysAspAlaGluIleLeuTrpGlyH	134
117	GTACATTTATTCATTCGCAAAAGATGCGAGATTTATTTATGCGCA	400

```

134 lathrhtlysglgluileuthrthrmetilegluarglyglutrasp 150
151 gluleuleuargyallysvallyspgoglyasphepheyvalpr 167
451 gacactctggccgctgtaaaaggccgggggattttttctatgctcc 500
167 oserylythrvalhlsalalleglylsglylleualaaleugluthrg 184
501 AACGGTACTGCTTCATCGCATTCGAAAAGAAATCTGCTTGGAGACGC 550
184 lnglnasnsesrpfthrthrtyrargleutyasptryasparglyasp 200
551 AGCAGAACTCAGACACACCTACAGATTATATGATATGACGAAAGAT 600
201 Alaggluglylyseuarggluleuhtleuhtlyseuhtleuhtleuht 217
601 GCAGAAAGCAGACGCGCGAGCTTCTATCTGAAAAGACACATTGAAGT 650
217 egluvalproserilleprogluarhishthrvalhshisgluinileg 234
651 AGAGGTCCGCTGATTCAGAAAGGCAATCAGCTTCAACGTAACAAATG 700
234 luaspheuleuthrthrthrleuhtleuhtleuhtleuhtleuhtleuht 250
701 AGGATTTGCTTACACGACATGATGAAATGCGCTTACTTTGCGTGGGG 750
251 lyastrpansleuhtleuhtleuhtleuhtleuhtleuhtleuhtleuht 267
751 AAATGGAATTCATCAGATCAGACACATTAAGCAACAAAACCAATCTCT 800
267 uleuhtleuhtleuhtleuhtleuhtleuhtleuhtleuhtleuhtleuht 284
801 TCTATCAGTGTATTCAGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 850
284 altyrprophelylsglyasphepheyvalprtyrtyrtyrtyrtyrtyr 300
851 TCTATCTTCTTCAAAAAGGAGATCATATCTGCTGCTTACGCTTCTGGA 900
301 Gluphelylsglylsglylsglylsglylsglylsglylsglylsglyl 315
901 GAATTTAACTCGAAGATATGCAAGATATGCTGCTGCTGCTGCTGCTG 945

```

seq_name: gp_ba2:BSUB0007

seq_documentation_block:

LOCUS BSUB0007 216750 bp DNA BCT 26-NOV-1997
DEFINITION Bacillus subtilis complete genome (section 7 of 21); from 1194391 to 1411140.
ACCESSION Z99110.1 AL009126
VERSION Z99110.1 GI:2633472

KEYWORDS
SOURCE
ORGANISM

Bacillus subtilis.
Bacteria; Firmicutes; Bacillus/Clostridium group;
Bacillus subtilis.
1 (bases 1 to 216750)

REFERENCE
AUTHORS
Kunst, F., Ogasawara, N., Moszer, I., Albertini, A.M., Alloni, G., Azevedo, V., Bertero, M.G., Bessieres, P., Bolotin, A., Borcherdt, S., Boriss, R., Boursier, L., Bruns, A., Braun, M., Brignell, S.C., Brown, S., Brouillet, S., Brusch, C.V., Caldwell, B., Capuano, V., Carter, N.M., Choi, S.K., Codani, J.J., Connerion, I.F., Cummings, N.J., Daniel, R.A., Denzot, F., Devine, K.M., Dusterhoft, A., Ehrlich, S.D., Emerson, P.T., Entian, K.D., Errington, J., Fabret, C., Ferrari, E., Foulger, D., Fritz, C., Fujita, M., Fujita, Y., Fuma, S., Galizzi, A., Gallener, N., Ghm, S.Y., Glaser, P., Gottreau, A., Golyzhnly, E.J., Grenat, G., Guisepi, G., Guy, B.J., Haga, K., Halech, J., Harwood, C.R., Henaut, A., Hilbert, H., Holsappel, S., Hosono, S., Hullo, M.F., Itaya, M., Jones, L., Joris, B., Karamata, D., Kasahara, Y., Kleier, Blanchard, M., Klein, C., Kobayashi, Y., Koeltter, P., Koningstein, G., Krogh, S., Kumano, M., Kurita, K., Lapidus, A., Lardinois, S., Lauber, J., Lazarevic, V., Lee, S.M., Levine, A., Liu, H.,

TITLE
JOURNAL
MEDLINE
REFERENCE
AUTHORS
TITLE
JOURNAL
The complete sequence of the gram-positive bacterium
Bacillus subtilis
Nature 390 (6657): 249-256 (1997)
98044033
2 (bases 1 to 216750)
Kunst, F., Ogasawara, N., Yoshikawa, H. and Danchin, A.
Direct Submission
Submitted (18-NOV-1997) I. Moszer, A. Danchin, Institut Pasteur.
Regulation de l'Expression Genetique, 28 rue du Docteur Roux, 75724
Paris Cedex 15, FRANCE. E-mail: moszer@pasteur.fr,
adanchin@pasteur.fr Phone: +33 (0)1 45 68 84 41, Fax: +33 (0)1 45
68 89 48

FEATURES
source

gene

CDS

gene
CDS

Location/Qualifiers
1. 216750
/organism="Bacillus subtilis"
/strain="168"
/db_xref="taxon:1423"
116. 1156
/gene="argC"
116. 1158
/gene="argC"
/EC_number="1.2.1.38"
/function="arginine biosynthesis"
/codon_start=1
/transl_table=11
/product="N-acetylglutamate gamma-semialdehyde
dehydrogenase"
/protein_id="CAB12976.1"
/db_xref="GI:2633473"
/db_xref="SWISS-PROT:P23715"
/translation="MKIGIVGATGCGTETLRLSHHPAEKCIYSSSGEGNYSEG
YPLHTGLADQOLKPIDMNTIKHEIDIMPLAEPGVSEETPKLADAGTVIDLSDLR
IKPEAEKMYKRTAPKAVIOAVYGLAELNLOIQOAKLIANGCEPPTAVILGLAP
LAOKKILDESFEIVDAKTGVSGAKRSGKSGTHESELDNRFKIYKNEHOPPEIRAL
NEMQPGIGPTTFSAHLVPMRTGMTATMTYRITCDLTDADLIDLYSEFYODSYFVVRP
KGOYRPOKREYVGSNFCIDIAVTLDERNTRTIYSVINDLKKGAAGAVONFNLMGNWE
ETGLTTPPIYP"
1176. 2396
/gene="argD"
1176. 2396
/gene="argD"
/EC_number="2.3.1.35"
/EC_number="2.3.1.1"
/note="alternate gene name: argA, argE, ornithine
acetyltransferase"
/codon_start=1
/transl_table=11
/product="amino-acid acetyltransferase"
/protein_id="CAB12977.1"
/db_xref="GI:2633474"
/db_xref="SWISS-PROT:P36843"
/translation="MTQLSEDOIVKVTQVSSPKGQKANGVHCGLRYSKKDLGVITSE
TPASAAVYVQSHFQAPLKVTDLSLKHGPTLKAIVNSALINACTGDEGLDAYTMR
ESFASQIGIPELVAVSSTVGIEHLDMERIHAGIELLKEPAGSGDEEALITVDY
IKQCYELAGKGTAVTIGARKSGSMIHNNATMLGFTYDAAIEKALOKLRETVD
VSFMQIVDETSTNDMVLVMANACENCLIEDHPDMVTFKFAALLCEIDAKIAR
DGEGATRLIEAVQAGAKNNINDANVIANKIYGSNVLVYAVYGTDAWMGRIGAIHSAA

```

gene
/translation="MKPTVTSISILVIGSGPILITIOQAAREPDSCTGGCIATKEGTRY
LTVYDIKINASYR"
2411..3187
/gene="argB"
2411..3187
/gene="argB"
/EC_number="2.7.2.8"
/function="arginine biosynthesis"
/codon_start=1
/transl_table=11
/product="N-acetylglutamate 5-phosphotransferase"
/protein_id="CAB12978.1"
/db_xref="GI:2633475"
/db_xref="SWISS-PROT:P36840"
/translation="MKRTVFKCGSVIRLESEFYONIKELRASGMKLAIVHGSGPE
ITNMLKRNITKEFSGGOKRTKPYLEVAEYVSGSYNKKFPFAELAKHGLRAAGISGK
DGLLEADYLDPEYGEVGEIKKYDASVMVALMENGIIPIVIAPIASMTSDCKTLNVAD
LAASVAGALEADKIMETVTDVGDIMKEKORLDTLPKEIOMLIKOEVIITGMIPKVN
ALSALSDQVSEVMYVNGKSPFADOTFOGTRIVAKKAVS"
3184..4341
/gene="argD"
3184..4341
/gene="argD"
/EC_number="2.6.1.11"
/function="arginine biosynthesis"
/codon_start=1
/transl_table=11
/product="N-acetylornithine aminotransferase"
/protein_id="CAB12979.1"
/db_xref="GI:2633476"
/db_xref="SWISS-PROT:P36839"
/translation="MSLFFQYGRWDIDIKKAKGTVEEDNGKTYLDFIOGIASVNI
HCHAEATVAVKROLDVWVHNSLRFONSLOPDAOKLAGHSGADLPFGNSGGAGCA
KLAARAKGKRTKRTPIOSFGRPYAGMAACDRIKTCGPMIGSPHYIPIYDNPAP
KALGEGCIIAAMLETVGSEGVNPARASERISAVOSCKRQMLIIIDETQITGRTG
KCFRHEHGLSDIITVAKGNGEPFGVAVIGKRODEATPDSHGTTGEMNLAAV
VNAIIOIVFQDPFOEADKGAFLKEIDELKSPVKOIRGGLMIGIECDGPVADI
IAELDTISLIVIPAGPERDSAARPLVTYKDEIAVASKLOALAHHSAAVQ"
4412..5473
/gene="carA"
4412..5473
/gene="carA"
/EC_number="6.3.5.5"
/function="arginine biosynthesis"
/note="alternate gene name: cpa"
/codon_start=1
/transl_table=11
/product="carbamoyl-phosphate transferase-arginine
(subunit A)"
/protein_id="CAB12980.1"
/db_xref="GI:2633477"
/db_xref="SWISS-PROT:P36838"
/translation="MEGYLVLEDGTSFSGELDHENCTGRAVFTGNTGYQEVLDPS
YKGOIIVFTYPLIGNVGINEDFESKRFQVAAAVVACHFSHYEAVYSLKEVLDKW
NIPILTHDTRAVVKKIRANGTGMATVATASGAELALQEPNVAEQASQEIETGDK
NKHILHIDBGYKSIASISLVKGCQKTVVAPPOGEAVYNNKPDGIVTSMNGPDKIO
PLVIGIKSISREFTLGICLGHOLALAFGNTKLPFGRRGHNHPVTDKTRKRVMT
SONHGSYVDEOSINEEELTIFPHNYNITSVEGLAKHKLPLVMSVQFHEAHFGAAEEM
ITDDYIKVITPARREIANA"
5466..8549
/gene="carB"
5466..8549
/gene="carB"
/EC_number="6.3.5.5"
/function="arginine biosynthesis"
/note="alternate gene name: cpa"
/codon_start=1
/transl_table=11
/product="carbamoyl-phosphate transferase-arginine
(subunit B)"
/protein_id="CAB12981.1"
/db_xref="GI:2633478"
/db_xref="SWISS-PROT:P318185"

```

```

/transl_table=11
/product="ornithine carbamoyltransferase"
/protein_id="CAB12982.1"
/gene="argF"
8537..9510
/gene="argF"
8537..9496
/gene="argF"
/EC_number="2.1.3.3"
/function="arginine biosynthesis"
/codon_start=1
/transl_table=11
/product="ornithine carbamoyltransferase"
/protein_id="CAB12982.1"

alignment_scores:
Quality: 1692.00 Length: 315
Percent Similarity: 100.000 Gaps: 0
Percent Identity: 100.000

alignment_block:
US-09-462-846-2 x BSUB0007
Align seg 1/1 to: BSUB0007 from: 1 to: 216750

1 MetThrThrGluProLeuPhePheLeuSerProValPheLeuGluArgIleTr 17
|||||
79774 ATGACGACATGAAACCGTTATTCTTCAAGCCTGTTTCAGAAAGAAATTTG 79823
17 pgiyglutThrAlaLeuAlaAspPheGlyTyrThrIleProSerGlnArgT 34
|||||
79824 GGGCGGAGCCGCTTTCAGCTGATTTGGCATACCATTCGTCACACAGAA 79873
34 hciylglucystrPalaPheAlaAlaHisGlnAsnGlyGlnSerValVal 50
|||||
79874 CAGGGAGTGTCTGGCTTTTGGCCGCATCAAAATGTCAAAGCGTTGTT 79923
51 GlnAsnGlyMetTyrLysGlyPheThrLeuSerGluLeuTrpGlnIshI 67
|||||
79924 CAAGAAGCAATGATGATGAAGGGTTTCAAGCTCAGCAGCAATATATGGAAATCA 79973
67 sarGHisLeuPheGluGlnLeuGluGlyAspArgPheProLeuLeuTrp 84
|||||
79974 CAAACATTTATTGGACACTTGGAAGGGAGCGTTCCCTCTGCTTCA 80023
84 ylleleuAspPalaAspGlnAspLeuSerValGlnValHisProAsnAsp 100
|||||
80024 AAATATATTAGTCTGACACGAGACTTATGTCAGTGACATCCAAATGAT 80073
101 GluTyrAlaAsnIleHisGluAsnGlyLeuLeuGlyLysThrGluCysTr 117
|||||
80074 GAATATGCCACATACATGAAAAAGCGTGAGCTTGAAAAAGATGATCTG 80123
117 pTyrIleIleAspCysGlnLysAspAlaGluIleIleTyrGlyHisAsn 134
|||||
80124 GTACATTATTGATTCGCCAAAGATGCCAGATTTATTATGCGCCACATG 80173
134 lahrhrlysgluGluLeuThrThrMetIleGluArgGlyGluTrpAsp 150
|||||

```

```

80174 CAAACAAGAGAGAGACTACTACATGATAGACCTGGAGAAATGGGAT 80223
151 G|U|L|E|U|A|A|A|G|A|G|V|A|L|Y|S|V|A|L|Y|S|P|R|G|I|A|S|P|H|E|T|Y|V|A|P|R 167
80224 GAGCTCTTCGCGCGTGAAGGTAAAGCCGGGGGATTTTCTATGTCGCC 80273
167 osergylthrvalhisalalleglylsyglytleuvalaileucjuphrg 184
80274 MAGCGGTACTGTCATGCGATTGGAAGAAATCTTCTTTGGAGAGCG 80323
184 lnglaasenseraspthrthrytrargleutyfaspfyrasparglyasp 200
80324 ACCAGAACTCAGACACACACTACAGATTATATGATTAGACCGAAAAAGAT 80373
201 Alagluglylsaleuarglulubuhlsaleuylserserilegluvalil 217
80374 CGAGAGGCGACACTCGCGACCTCATCTGAAAAAGACATTGAAGTCAT 80423
217 egluvalproserileprogluarghisthrvalhisglucjulinleg 234
80424 AGAGGTCCGCTGATTCCAGAACGCGATACAGTTCACCATGACAAATG 80473
234 luaspleuueuthrthrthreulleuglysalatypheservalgly 250
80474 AGGATTTCGTTCAACGACATTGATTGAGTGGCTTTCGTTGCGTGGG 80523
251 lvsitpashleuSerGlySerAlaserleuylsnglinlvsprophela 267
80524 AATGGAAGTCTTCAGAGTACAGACAGCTTAAAGCAAGCAAAACCATTCCT 80573
267 ubeuileservalilegluglylgluylargmetileserglylunryr 284
80574 TCTTATCACTGATGATTGAAGGAGCGCGCTATGATCTCTGTGAGTATG 80623
284 altyrprophelylsyglyaspHismetleuleuprotyrGlyleugly 300
80624 TCTATCCTTCAAAAAAGAGATCATATGTTGCTGCCCTTACGGCTTGA 80673
301 GuipheylsleuglytyrAlaglysalilevalserHisleu 315
80674 GAATTTAACTGAAAGATATGCAAGATGATCGTCCCATCTG 80718
seq_name: gb_bai:AB011838
seq_documentation_block:
LOCUS AB011838 15012 bp DNA BCT 24-FEB-1999
DEFINITION Bacillus halodurans C-125 genomic DNA, 6A fragment, clone ALBAC004.
ACCESSION AB011838
VERSION AB011838.1 GI:4512380
KEYWORDS
SOURCE
ORGANISM
Bacillus halodurans (strain:C-125) DNA, clone_lhb:lambda no.4
clone:ALBAC004.
Bacteria; Firmicutes; Bacillus/Clostridium group; Bacillaceae;
Bacillus.
REFERENCE
1 (stiles)
Takami,H., Nakasone,K., Ogasawara,N., Hirama,C., Nakamura,Y.,
Masui,N., Fujii,F., Takaki,Y., Inoue,A. and Horikoshi,I.
Sequencing of three lambda clones from the genome of alkaliphilic
Bacillus sp. strain C-125
Extremophiles 3 (1), 29-34 (1999)
99184646
2 (bases 1 to 15012)
Takami,H. and Inoue,A.
Direct Submission
Submitted (04-MAR-1998) to the DDBJ/EMBL/GenBank databases. Hideo
Takami, Japan Marine Science and Technology Center, Deep-sea
Microorganisms Research Group; 2-15 Natsushima, Yokosuka, Kanagawa
237, Japan (E-mail:takami@jamstec.go.jp. Tel:81-468-67-3895,
Fax:81-468-66-6364)
Sequence updated (28-Apr-1998).
Location/Qualifiers
1. .15012

```

```

/organism="Bacillus halodurans"
/strain="C-125"
/db_xref="taxon:86665"
/clone="ALBAC004"
/clone_lhb="lambda no.4"
/map="GA"
/note="alkaliphilic Bacillus"
4. .9
/note="SD sequence"
21. .1421
/gene="imgE"
21. .1421
/gene="imgE"
/note="similar to B.subtilis imgE gene(74%-identity)"
/codon_start=1
/transl_table=11
/protein_id="BA075344.1"
/translacion="MLEQADVYLHHEITSEEFSTAHYVLIDTIGCGTILANPECT
KHGPIVPGTIVRGTRVPGTSYLDNPGAPICGMVWVLDYNDPKLAVEGHSDN
LGLIADADTISDENLSKKEPLKYKDYLEMKAHEIOGYLALENSLRVLGYLIF
VKATTAIVYTKLGGDROETINLSNKNIDNSLRTKQAPNTGSKSPAGDAYSRA
VRAMWAIRKESKGIQRLSAPGFDVDFPNKQKLVLPDSYVMEVLEVSTPAE
FHAQTAERAEIIEHPEYGRHDEIDHTTSTESAIRIDKQPLYNPMDRCHLOYL
TAGLNGALVAEDYEDEAAADPRIDLRLDKKVIENKDYTDQYLDPPKRSIANKVOY
HFQDGTSTANIEKQYPLGHRFRREAVPRIVEKYANLGTTHYAKQLEIAVSVDYD
RLAAMNVEIETLFFVK"
1425. .1430
/note="SD sequence"
1439. .2341
/gene="yqjO"
1439. .2341
/gene="yqjO"
/note="similar to B.subtilis yqjO gene(68-77% identity)"
/codon_start=1
/transl_table=11
/protein_id="BA075345.1"
/db_xref="GI:4512382"
/translacion="MAWVEQPSQKELADRRQLMKKEALLIQIFGANDMAALVAKK
AGFSALYLSGDVATASRGLEPDLGIIVSTEVADAKDLVATNLPLVLVDIDFGGVLN
VARTAOEMLSEDAVNAVAQVIEDQDLPKRCGLNGQLSKSEMDOKLOALRKVAPLTVI
ARTDARANEGILNGALIERANVYIEAGADALFPEALQSAEERFLVAENVSPFLANNTEF
GKPLMTAGGLQNAQFQWVLYPVTSLRVAKAKVERIFOLIKDEGTQGEAGIEDMQRKE
LYETISYDPEALDKNIATVLGE"
1856. .1863
/gene="yqjO"
/note="SD sequence"
2939. .3196
/gene="ydbN"
2939. .3196
/gene="ydbN"
/note="ydbN"
/note="similar to B.subtilis ydbN gene(59%-identity)"
/codon_start=1
/transl_table=11
/protein_id="BAA75346.1"
/db_xref="GI:4512383"
/translacion="MKTELMADPADYPSIDKLVIHIISSNNSYSHDVMOKIOTSLNKSII
DIRKLIDYNRVAKALYHKKADCNKEDFGLTVRECOQV"
3446. .3451
/note="SD sequence"
3458. .5105
/gene="ydhO"
3458. .3787
/gene="ydhO"
/note="ydhO"
/note="similar to B.subtilis ydhO gene(69%-identity)"
/codon_start=1
/transl_table=11
/protein_id="BAA75347.1"
/db_xref="GI:4512384"
/translacion="MEALDFDQISSDQIFPMIILISGNGRSKRILOALREYREGKAPA
NDLVEAEKDLNVAHEIHPKLVQGEATGKRTLSLLNHAEDHLMSTLTJMDVYKELL
EIFKMK"
3799. .3804

```

RBS

RBS /gene="ydhO"
/notes="SD sequence"
4222. 4226
/gene="ydhO"
/notes="SD sequence"
4239. 5105
/gene="ydhO"
/notes="similar to B.subtilis ydhO gene(81%-identity)"
/codon_start=1
/transl_table=11
/protein_id="BAA75348.1"
/db_xref="GI:4512385"
/translation="MNTNPLAEIYRIRVORNTITMPCGPPAPAYAKSFALLIPAVLT
LYFLANLIVFSQLEHTNMADYVITQAVQAPLVSGLGIIPLTILALIFLQILMFGLNG
QVILNSVMPILNLTLEMLATYNGEPVNNIIISKPFITFTFVGNGGTMLAVPAFI
LLEKMSQQLKQVKGGLGIPGLFVNNEPILEGDLVNNPLIIVPMLISMLVITLVYS
MAGLIPVPTGVAVFWIVPFPIINGIMATNSIAGLLQILVNIIMVEMWYPLPKSMDFRE
NKKEDDEKSSA"
5104. 5111
/notes="SD sequence"
5113. 6474
/gene="ydhP"
5113. 6474
/gene="ydhP"
/notes="similar to B.subtilis ydhP gene(80%-identity)"
/codon_start=1
/transl_table=11
/protein_id="BAA75349.1"
/db_xref="GI:4512386"
/translation="MPCMTKDTVTYRPRPEGFWMGSAATQTEGATHRDGKGIIMPHD
WYDESRPFQVCGPODTSOPYERYKEDIDLMKEVGHNSFRFSISMSRLIDPGEGEY
PKGVYKNVNIIDELGOGIEPFWNLPHFDMPELQKIGMANREYVDLFAVATGEEN
RDYRKYKNVINEPIIVPEGGYLIDPHYENVEDPERKAIQVGYHTVLSAAATRAYKO
LNGOGIGIILMLITSYSPRSNHPADIKAAKILADAPFNSEPLSPKSGFPEELVQIILK
EESRPMILIEEPLLEPIIRHOTVLDLGAVTYQPRRYAKETLIPMDAPMDEYFPHYIYM
PKRKMYHHGNETIYERKGIYDILMLKHYGNICPISBNGMYVEDEEYFPRDENVYID
DIRLEIQLQHLKHAIAIQEGSNVKGTHLMTFDMNSMANATKRLMVGVCQFROENG
TDN"
RBS 6505. 6509
/notes="SD sequence"
6513. 7313
/gene="ydhQ"
6513. 7313
/gene="ydhQ"
/notes="similar to B.subtilis ydhQ gene(53%-identity)"
/codon_start=1
/transl_table=11
/protein_id="BAA75350.1"
/db_xref="GI:4512387"
/translation="MDPKRWSYPDHLPLKINQSLQKVGDAKKYEALISMEIKRKIK
ENYVADQPIPEBISLSKFCGSRMTMKRALDILVMEGLLYRRQHGFIYQSAIHRS
RVNVLNVEYGSJNLKDKQVSKYQIYNVEFAEEVAALIDKQTPVYHILRLKLV
DGEYVLYETKMPITPLIGINDENVLHGSISYLLDKGLIAGHSKRIACKAEALTELE
EHLHCEADDPLEIHEVHGLNNGVFEYSFRRHYDKFEVTTIMKRS"
7663. 7667
/notes="SD sequence"
7672. 8619
/gene="ydhS"
7672. 8619
/gene="ydhS"
/EC_number="5.3.1.8"
/notes="similar to B.subtilis ydhS gene(53-69% identity)"
/codon_start=1
/transl_table=11
/product="mannose-6 phosphate isomerase"
/protein_id="BAA75351.1"
/db_xref="GI:4512388"
/translation="MLQPIFLTPYFKERIMGTLTKQEGYNIIPSDTKGECMAISAHH
NGSSIVANGFEICKSLAEILKRPOLRGNPSSVYPLIYLIIDANDLSVYGHDDSS
ANSHEMSGLTTCWTLIDCKDAKMLIHHAKTREFLEKICEDGAMDDLTKRYATKIK
GDFEYVSGITIALCEGTLVLTQSSDPTIRVYDIDRDEANLEFELHELEADIVTTL
VLTQDPVCKVETIKQDAVVTVEFTEFSVHKMDHGQSFSSDDRYLVAISILEGTGTE
RLTVQGRGSRLEKGAILIIVGFNDPNTIEGQCOOLVSHP"

```

RBS      8061.. 8063
          /gene="ydh8"
          /note="5D sequence"
          /complement(8719..10722)
          /gene="mcnpA"
          complement(8719..10722)

CDS
          /gene="mcnpA"
          /note="Similar to B.subtilis mcnpA gene(28k-identity)"
          /codon_start=1
          /transl_table=1
          /product="methyl-accepting chemotaxis protein"
          /protein_id="BAA5352.1"
          /db_xref="GI:4512389"
          /translation="MIRKLSFFRRKSIINLKFIVYFLVGLIPLIIMSVMYMOSSSENIIV
          AKERAMSESQTVSTVDGMEOMLEKRILDEKIVASFSALSDSEKSLQIDMNVINQDD
          SEYEVFLTGADGIVQVATNEHIGVLDLSIREFQMGAGQPTISVLTISATNARIV
          VAVPVLSSGSEGLGVMSASINFEOLIAQIHDDQFNOHEVFIIDQVTOAHNDELF
          IGIIVESGIESDVLSPATGKESGSGSIRVDSGEMLSIVPFIKACGYGLYLLPSS
          SLVATYDHIDQITIMAVPFLVITVPAIVARSMSRPRIITTHSVOTIAGDGLTSD
          LSQIKDDQDGIILANNNOAMFNLEKVIHOVGSADRVASQOOLASSEVSRSEDO
          IISQIIOVSGSEKSGSVNENSTLIEFVSSVQOLAASVQITSEANSTIKARGGG
          STENVEYPMKTIHESVDYQVOTIKMAERSEKITNLIDVTSITIAQTNLAINAITEK
          ARACHGKFAVAVVDVPRKLAESGHSQOISIVYELQIKENDPTSTAMQNMEEVND
          GLTVANERTKTHQETIOSTNOVAPQIEMAAITSEETSAVEQVSSAPFIEAVTAEKTS
          DSAQVVAASEQQLAIMEVYSSAQVLSQAMATURLKSEKY"
          complement(10733..10733)

```

alignment_scores:	Quality:	988.50	Length:	316
Ratio:	3.831	Gaps:	3	
Percent Similarity:	81.646	Percent Identity:	57.595	
alignment_block:				
US-09-462-846-2 x AB011838 ..				
Align seg 1/1	to: AB011838	from: 1	to: 15012	
4	GIUPRDEUpherPhelysProvalPheylsGluatgllletrpGlyglYth	20		
7678	CACCGATTTTTTTTAAACCCGTGGTTTAAAGGACCGATTTGGGGGGGAC	7777		
20	lAlaleu...AlaAprheglYtYthrlleProserGlnatgYthrlcylg	36		
7728	TCTGCTTAGCAGAGATTTGGCTACATATTTCCATCCGATTAACGGGAG	7777		
36	lucyETPRAlarPhAlAlAniSglAsnGlylInserYalAlGlnsn	52		
7778	AGCTGGGCGCATTTTGGCTGCATCCGACGGCGCTCTATTTGTTCTGAT	7827		
53	GlyMetYtYlsglyRhetrlleuSerGlnleuTrrpLinhisAlArgH	69		
7828	GGCGAATTTATTTGGAATCATTTAGCAGACCTTTGGAAGGACACCGCA	7877		
69	steuPheglYcInleuGluclYlAsparGtrPheProleuLeuThrLyslle	86		
7878	ACTATTGGAAATCCCTTCATCAGACGCGTTCCGTTATTTAAAGAAAATTT	7927		
86	eulAsPAlAsPglAsPleuSerYalGlnValhisProAsnAspLutYr	102		
7928	TAGACGCCAATATGATTTATCCGTGACGGTTGACCCCGATACACATAT	7977		
103	AlaAsnllleHlsGluAsnGlylLeuGlyLysThrGluCystrpYrll	119		
7978	GCCAAACACCATGAAATGGGGATTTGGAAGAAGCAGATGCTGGTACAT	8027		
119	elIAspCystrGlnlyAspAlaGluIleleYrGlyHlsAsnAlaThrT	136		
8028	CCCTATTTGTAAAGAAAGTGCAGAGATATTTAGTCCACCTGCAAAA	8077		
136	hLYsGluGluLeuThrThrMetlleGluArgGlylUtrPAspLeu	152		
8078	CGAGGAAGAAATTTAGCAAAATGTTAAATATGGGCAATGGGATGATTTA	8127		


```

CDS                                     /gene="BH3778"
                                     complement(846..1301)
                                     /gene="BH3778"
                                     /note="BH3778
                                     unknown"
                                     /codon_start=1
                                     /transl_table=11
                                     /protein_id="BAB07497.1"
                                     /db_xref="GI:10176403"
                                     /translation="MRKGWIAACWALTCIYGCEREMPRQAOGLNERPAGOVYIGLGP
                                     GANDYNTYKRSIADAPRONSGSPRPAPYPTPTIADIDERTISITIDDEHGLRSSMW
                                     MRGSHAVYVNEFPFRMKKEORNKIMNDLREDMOQIRRYOYHLSQKG"
gene                                     complement(1382..1966)
CDS                                     /gene="tdk"
                                     complement(1382..1966)
                                     /gene="tdk"
                                     /EC_number="2.7.1.21"
                                     /note="BH3779"
                                     /codon_start=1
                                     /transl_table=11
                                     /product="thymidine kinase"
                                     /protein_id="BAB07498.1"
                                     /db_xref="GI:10176404"
                                     /translation="MOMIKEGMIEYIGSGMFSKSEELIRVRATYGRKLYQFKP
                                     AIDNRYSEGEVYSHNGNKVCALPEVEMSLIOWSEDDIVAIDVQFPEDEVIVVD
                                     YIADHGVYICAGLDODFRGEPERFLMAEYVTKLQALCPVCSGSPASRTORLID
                                     GPPASYHDEIILVGAAYEPRCRHCHOVYTKRSLPETSASSE"
gene                                     complement(2159..2404)
CDS                                     /gene="rpmc"
                                     complement(2159..2404)
                                     /gene="rpmc"
                                     /note="BH3780"
                                     /codon_start=1
                                     /transl_table=11
                                     /product="50S ribosomal protein L31"
                                     /protein_id="BAB07499.1"
                                     /db_xref="GI:10176405"
                                     /translation="MKPGIHPYKVVFPMDSTGFKLSGTSKTSNETIEMEDGNTYP
                                     LKVEISSDSHPFYTKRIADAGGRVDRFKKYNLK"
gene                                     complement(2516..3787)
CDS                                     /gene="rho"
                                     complement(2516..3787)
                                     /gene="rho"
                                     /note="BH3781"
                                     /codon_start=1
                                     /transl_table=11
                                     /product="transcriptional terminator"
                                     /protein_id="BAB07500.1"
                                     /db_xref="GI:10176406"
                                     /translation="MGVHLSELENMKRLRELYELARKYKISYYSKITKKELIFAILKQ
                                     AEDGILFMGVLEIIOSEGFLRPINYPSEEDYISASQIRREDLNGKVSQV
                                     RPKENRYHGLIHEAVNGEDPTSKERHPFALPLPEKRLNETTNGRISRI
                                     DMISVFGQGRGLIVAPKAGKSLIKAIANSIVHNPVELIIVLIDRPAVYNDI
                                     RSYDADYVSTPEPEVNHKVAELIEMARIVYERKKTIVILIMOSITPLAAYNIV
                                     PEGGRILSGSIDPAFAHPRKPRGGAARNEEGSUTLILATVYEGSSRMDVITYEFK
                                     GSTNMELHDKRLSERIRIFALDIRSSTRKEELMKLEKMAVARTMDSPEFY
                                     DHEIRKKEKTKTDEFEEAMKDMNGTKRK"
gene                                     /gene="BH3782"
                                     /gene="BH3782"
                                     /note="BH3782"
                                     /codon_start=1
                                     /transl_table=11
                                     /protein_id="BAB07501.1"
                                     /db_xref="GI:10176407"
                                     /translation="MNHRYFEGINEWEAIKITEVEVMEBHCLSIRTSITIIHGI
                                     IQYVNNFEGALSERVIGELGNHQAQSVLEOAPDRILYLMTRSGFFMCEPSTK
                                     RTVPDFARMTRD"
gene                                     complement(4164..5129)
CDS                                     /gene="BH3783"
                                     complement(4164..5129)

```

```

/gene="BH3783"
/codon_start=1
/transl_table=11
/product="glycerol-inducible protein"
/protein_id="BAB07502.1"
/db_xref="GI:10176408"
/translation="MERSLSMELVRYTEAALASARWNGRKNKEADRAATEARDVF
DTIPMKGTAVYEGEMDEAPMLYIGELGNGYSPRVAVADPLEGTNIVASGIMNLA
VIAVADGNLILHAPDMYEMKIVAPGVADIDAPYIDMLRAVAKKNADVGLVYV
TINKRDKLIDELROAGARKILIPDDVAAAINAPDQDGVLLGSGAPGVLA
VGKCLGELGLOGKLLPOTDEILARCKMAGIEDVKNYSEIDELVKGDCCIPAAVGVDS
ELKGVYRKESFATQSLVMRAKSGYRFPDGSKMKKRDLYIK"
gene                                     complement(5173..6459)
CDS                                     /gene="mut2"
                                     complement(5173..6459)
                                     /gene="mut2"
                                     /EC_number="2.5.1.7"
                                     /note="BH3784"
                                     /codon_start=1
                                     /transl_table=11
                                     /product="GDP-N-acetylglucosamine
                                     1-carboxyvinyltransferase"
                                     /protein_id="BAB07503.1"
                                     /db_xref="GI:10176409"

alignment_scores:
      Quality: 988.50      Length: 316
      Ratio: 3.831      Gaps: 3
Percent Similarity: 81.646      Percent Identity: 57.595

Align seg 1/1 to reverse of: AP001520 from: 1 to: 305153
US-09-462-846-2 x AP001520/rev ..

4  GlnProLeuPhePheLysProValPheLysGluArgIleTrpGlyGlyh 20
:::|||||:::  ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
155351 CAACCGATTTTAAACCTGCTGTTAAAGACGATTTGGGAGGAC 155302

20  rAlaLeu...AlaSPheGlyTyrThrIleProSerGlnArgThylGly 36
| |||  ::|||::| ||| ||| ||| ||| ||| ||| ||| ||| |||
155301 TCGGCTTAACGACGAGTTGGCTACATATTCATCCGATAAAGCGGAG 155252

36  LucySTPALaPheAlaAlaHisGlnAsnGlyGlnSerValValGlnAsn 52
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
155251 AGTGTGGGCAATTCCTGCTATCCGACGCGCTCTCTATGTGCTAAT 155202

53  GlyMetTyrLysGlyPheThrLeuSerGluLeuTyrPgluHisIleArgH 69
|||  :: |||  ::|||::| ||| ||| ||| ||| ||| ||| ||| |||
155201 GCGGAATTTATTTGGAAATATCATTAAGACGCTTTGGAAGGACCGCGCA 155152

69  sLeuPheGlyGlnLeuGluGlyAspArgPheProLeuLeuThrLysIle 86
||||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
155151 ACATATTGGAATTCCTCATCAGCGGTTCGCTATTATACGAAATTT 155102

86  euSPaLaSPGlnASPLeuSerValGlnValHisProAsnAspGluTyr 102
||||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
155101 TAGACGCCAATATGATTAACGTCGAGGTTCACCCGATGCTCATAT 155052

103  AlaSnIleHisGluAsnGlyGluLeuGlyLysThrGluCysTrpTyrI 119
||||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
155051 GCCAACAGCATGAAATGGGGAATAGCAAGACGAAATGCTGTACAT 155002

119  eLLeaSPCySGlnLysAspAlaGluIleIleTyrGlyHisAsnAlaThr 136
|:::|||||:::| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
155001 CCTTGATTTGAAGAAAGATGCAGAGATGATTAAGTCACCAATGCAAAA 154952

136  hTrLysGluGluLeuThrMetIleGluArgGlyGluTyrAspGluLeu 152
|:::|||||:::| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
154951 CGAAGAGAAATTAAGCAAAATGATTAAGATGGGATGGATGATTA 154902

153  LeuArgArgValLysValLysProGlyASPhePheTyrValProSerG 169
||||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||

```

```

154901 TTGCGAAGGTAAGATCAAGCCGCGAGATTTTATGTGCGCAGTGG 154852
169 YTHVAlHsAlAlleGlyslYleuAlaleuGlutHrGlInGla 186
154851 GAGGATCCAGCAGCTATGCGAGGAAAGCGCTCTTGAACACTCAGCAAA 154802
186 snserAptHrThrTYrArgLeuTYrAspTYrAspArgLysAspAlaGlu 202
154801 GCTCGATACGACGATACGCGGTGTATGACTACAGCCGCGAGATGAAGCG 154752
203 GlyLysLeuArgGluLeuHsLeuLysSerIleGluValIleLeuLys 219
154751 GGAATATTACGGAGCTTCATTATGAAAGCGCATACGTAACGACAGT 154702
219 L.....ProSerIleProGluArgHsThrValHsHsGluG 232
154701 GCCCATCAAGATCCAGACCTCTGTCCAAAGATGAAGATGAAGACCAAG 154652
232 InIleGluAspLeuLeuThrThrThrLeuIleGluCysAlaTYrPheSer 248
154651 CCGTT.....GTGACGACGTTGTGTAGAAAGGAGTCTTTTCC 154614
249 ValGlyLysThrPasnLeuSerGlySerAlaSerLeuLysGluGlnLysPr 265
154613 GTTCATTAATGGGACATTCACGAGACAGGCTTCTCTCTAGTACGATCG 154564
265 oPheLeuLeuIleSerValIleGluGluGluArgMetIleSerGlyG 282
154563 TTATTTACTGTAGATCATGTGAGAGACAGCTTACGTATCAAG 154514
282 LuTYrValTYrProPheLysGlyAspHsMetLeuLeuProTYrGly 298
154513 GCGAGACATCTCTTTGAAAGAGGCGCCACCTCTTAATCCAGTAGAGG 154464
299 leuGluGluPheLysLeuGluGlyTYrAlaGluCysIleValSerHs 314
154463 TTTAAATTTTAACTTGAAGCGCAATGCCAATGTATGCTTTCCAT 154416
seq_name: gb_pat1:AX006203

seq_documentation_block:
LOCUS AX006203 948 bp DNA PAT 24-NOV-2000
DEFINITION Sequence 4 from Patent WO9904016.
ACCESSION AX006203
VERSION AX006203.1 GI:9929072
KEYWORDS
SOURCE
ORGANISM Bacillus subtilis.
Bacteria: Firmicutes; Bacillus/Clostridium group;
Bacillus subtilis
Bacillus/Staphylococcus group; Bacillus.
REFERENCE
1 (bases 1 to 948)
AUTHORS Estell,D.A.
TITLE Proteases from gram-positive organisms
JOURNAL Patent: WO 9904016-A 4 28-JAN-1999;
GENENOR INTERNATIONAL B V (NL); ESTELL DAVID A (US)
FEATURES
source 1..948
/organism="Bacillus subtilis"
/db_xref="taxon:1423"
BASE COUNT 286 a 192 c 244 g 226 t
ORIGIN

```

```

alignment_scores:
Quality: 978.50 Length: 316
Ratio: 3.822 Gaps: 1
Percent Similarity: 81.013 Percent Identity: 56.646

```

```

Alignment_block:
US-09-462-846-2 x AX006203

```

```

Align seg 1/1 to: AX006203 from: 1 to: 948

```

```

1 MethTrhrGluProLeuPheLysProValPheLysGluArgIleTr 17
||||| ..... |||||||
1 ATGAGCGCATCCCGATTCTTAAACGCTGTGTAAAGAAAAATCTG 50
17 pGlyGlyThrAlaLeuAlaSP...PheGlyTYrThrIleProSerGln 33
||||| ..... |||||||
51 GGGCGGAACCCCTTACGAGATAGATTGGATGACAGATTCCCTCAAAAT 100
33 rGThrGlyLysTrpAlaPheAlaHsGlnGlnGlySerVal 49
||||| ..... |||||||
101 CAAGCGGGAATGCTGGCCATTCCCTCATCCAAAGAACGACGACACT 150
50 ValGlnAsnGlyMetTYrLysGlyPheThrLeuSerGluLeuTrpGluH 66
||||| ..... |||||||
151 GTTCAATATGGCCGTATTAAGGAAAGACATGTATCGATTTGGGAAGA 200
66 sHsArgHsIleuPheGlyGlnLeuGluGlyAspArgPheProLeuLeu 83
||||| ..... |||||||
201 GCACCGTAAGTATTCGGCGGTAGAGGGGATCGGTTCCGCTCTCTGA 250
83 hTrpLysIleLeuAspAlaSPGlnAspLeuSerValGlnValHsProAsn 99
||||| ..... |||||||
251 CAAGCTGCTGATGTAGAGGATACGTCATTAAGTTCACCTGAT 300
100 AspGluTYrAlaAsnIleHsGlnAsnGlyGluLeuGlyLysThrGlu 116
||||| ..... |||||||
301 GATTAATATGCGCGGAGAAAGCAAGAGGAGAACTCGCAAGACGGAATG 350
116 sTrpTYrIleIleAspCysGlnLysAspAlaGluIleIleTYrGlyHsA 133
||||| ..... |||||||
351 CTGTACTATTCACGCTGTAAGGAAAGCAAGAAATCATTTACGGCATTA 400
133 snAlaThrThrLysGluGluLeuThrThrMetIleGluValGlyGluTrp 149
||||| ..... |||||||
401 CGGCGCGTCAAAACCGAACTTCTCAGATGATCAACAGCGGTGACTGG 450
150 AspGluLeuLeuArgValLysValLysProGlyAspPheHsTYrVal 166
||||| ..... |||||||
451 GAGGCGTGTGCGAAGAAATCAAAATTAACCGGTCATTTCTATTATG 500
166 lProSerGlyThrValHsAlaIleGlyLysGlyIleLeuAlaLeuGlu 183
||||| ..... |||||||
501 GCCAGCGGGAAGCGTGCAGCGATGTGCAAGGGGCGCTGTTTGAAGA 550
183 hGlnGlnAsnSerAspThrThrTYrArgLeuTYrAspTYrAspArgLys 199
||||| ..... |||||||
551 CTCAGCAAAATTCAGATGCCACATACCGGGGTATGATATGACCGTCT 600
200 AspAlaGluGlyLysLeuArgGluLeuHsLeuLysSerIleGluVal 216
||||| ..... |||||||
601 GATGACCAAGGAAAGTCCGAGAGCTTCATTATTTGCCAAAGCGGTCAATGC 650
216 lIleGluValProSerIleProGluArgHsThrValHsHsGluGln 233
||||| ..... |||||||
651 CGCCAGCGTCCCATGTGCGACGGGTATATAGATGATCAAGCAATCAAA 700
233 lGluAspLeuLeuThrThrThrLeuIleGluCysAlaTYrPheSerVal 249
||||| ..... |||||||
701 GAAAGAAATATACCATTAATAACATTGTCCAGGGAATATTTTGGCTT 750
250 GlyLysTrpPasnLeuSerGlySerAlaSerLeuLysGlnGlnLysProP 266
||||| ..... |||||||
751 TATTAATGGACATCAATGGCGAGCTTGAATAGCTTCAGATGAATCTT 800
266 eLeuLeuIleSerValIleGluGluGluArgMetIleSerGlyLys 283
||||| ..... |||||||
801 TCTGATTTCAGCGCTGATAGAGAGACGCTTGTCAATATGACAGACA 850
283 YrValTYrProPheLysLysGlyAspHsMetLeuLeuProTYrGlyLeu 299
||||| ..... |||||||
851 AAACATGTCGCTCAAAAAGGTGATCATTTATTTGCCGGGTCAATG 900
300 GlyLysPheLysLeuGluGlyTYrAlaGluCysIleValSerHsLeu 315

```

```

seq_name: gb_ba2:BSU02562
seq documentation block:
LOCUS      BSU02562          3962 bp      DNA          BCT          03-JUN-1995
DEFINITION Bacillus subtilis N-acetylglucosaminidase (lytD) gene, complete
ACCESSION  U02562
VERSION     U02562.1 GI:476091
KEYWORDS
SOURCE      Bacillus subtilis.
            Bacillus subtilis.
            Bacteria; Firmicutes; Bacillus/Clostridium group;
            Bacillus/staphylococcus group; Bacillus.
REFERENCE   1 (bases 1 to 3962)
            Margot, P., Maue, C. and Karamata, D.
            The gene of the N-acetylglucosaminidase, a Bacillus subtilis 168
            cell wall hydrolase not involved in vegetative cell autolysis
            Mol. Microbiol. 12 (4), 535-545 (1994)
JOURNAL     95020588
MEDLINE     95020588
AUTHORS     Vladimir V. Lazarevic, TITLE Direct Submission.
JOURNAL     Submitted (18-OCT-1993) Vladimir V. Lazarevic, Institut de
            Genetique et de Biologie Microbiennes, Rue Cesar-Roux 19, Lausanne,
            1005, Switzerland
FEATURES
            source          1..3962
                        /organism="Bacillus subtilis"
                        /strain="168"
                        /db_xref="taxon:1423"
                        /map="310 degrees"
            -35_signal      3..8
                        /note="sigma"
            -10_signal      30..35
                        /note="sigma"
            misc_feature    41
                        /note="transcription start site"
                        /evidence=experimental
            RBS             169..177
                        /evidence=experimental
            CDS             186..1136
                        /note="orfX"
                        /codon_start=1
                        /transl_table=11
                        /product="unknown"
                        /protein_id="AA67856.1"
                        /translation="MQSPFLTPVPEKRIWGTALRDREGYSPSESTGECMAISAH
                        PKGSIYANGPYKGLIELMEHREVFEGVGDRPPLTKLLDKEDTSLKHPDDY
                        YAGNEEGELGTEKCIWIIDCKENAEIYGHTARSKTELVTWINGDWEGLRKIK
                        PGDFYVPGSTLALCGALVETQONSATYRVYDYLDSNGSPRLPAKAVNAA
                        TPVHDY IDESTSRKGIITKTFVGEYEVKMDINEAMADESFLICSVESG
                        GLKYEKTCPLKGDHFLPAMPDFIKTGCTLIYSHI"
            terminator      1135..1163
            -35_signal      1193..1197
            -10_signal      1213..1219
            misc_feature    1227
                        /note="sigma"
                        /note="transcription start site"
                        /evidence=experimental
            RBS             1254..1260
            gene            1265..3907
                        /gene="lytD"
                        /gene="lytD"
                        /EC_number="3.2.1.30"
                        /codon_start=1
                        /transl_table=11
                        /function="hydrolyses bond between the
                        N-acetylglucosaminyl and the N-acetyluranyl residues in

```

```

the glycan chain"
/product="N-acetylglucosaminidase"
/protein_id="AA67857.1"
/db_xref="GI:476093"
/translation="MKRRLIAPMLISAASIAFPAMSGSAQNAAYTDYSLYKVEPNTF
STESASOAVAKLEKDTGMDASVOASTTTTQYISAGHISEEAKIISGLAKQTSI
TGTSSPVGSKOPYATISSGATSGEKNATILAKIDETGAVAKVAGAPYMANVOT
SDIAPETVKLIQSLAKQNGIKSTYOPITHTVSVTIIGSTIVGSRBAOLIKNAFOT
ESGLASIKETNGQAYTETTAISGEANAKTILQIKOSTGITGSKINOKTVE
SYNNOAYKEKSTIKDAISOIKRNPVSGSYOOGKSTYTNMGKITKQOLKIDT
PEKKKMYTSSSVKTTTSAAYOTITAKIIGBOAKKAAPFAOKRYAKATK
AENQQLISEISDQAYTKGLNLTAKNQLSASAKSKYKOIATPEKTTESLIDTKY
NOALFEKSNHISVASOKTGOFAASTQITTEALISSEIDRVLTFFKQHLAVTSK
TGOATVTOYKILVTTOLSKTALNNGLTLYLKSSTVPSYTKTSNTLKISVNEQFOND
TAAASTRKOLQYGTSSIVKIKNGPQIMKTNYLSLIDKQOMTVSPQTDAAVVS
LVIYIATSTVYADVLNIRSTPEVSPNVIQGFPRGDKVVIQIINGMAKINLGMWNA
SSDEVVOYVDPNFSDSKSYFOPKLQSTAGLSYEVNOKVLGAKGILNGRAKAFID
AANQSIINELYLISHALETNGTSLALNGLTLYNGKTVNMVIGAYDSNPYNGAKY
AVEQGMFPEAIIIGAKFISSSYIHMTAVYNOTLYKMRWSATATHOYATDIGMAVKQ
VRMTSLYSLDGYTLFYDVEYR"
3924..3962
terminator 1358 a 871 c 875 g 858 t
BASE COUNT
ORIGIN
Alignment scores: 978.50 Length: 316
Ratio: 3.82 Gaps: 1
Percent similarity: 81.013 Percent identity: 56.646
alignment block:
US-09-462-846-2 x BSU02562 ..
Align seg 1/1 to: BSU02562 from: 1 to: 3962
1 MethrThrGluPheProLeuPheProValPheLysGluArgIleu 17
|||||
186 ANGAGCATACCGCATTTTCACCGCTGTCTTAAAGAAATAATCTG 235
|||||
17 pGlyThrAlaLeuAlaSP...PheGlyThrIleProSerGln 33
|||||
236 GGGCGGAACCGCTTACGAGATGATTTGATACGATATCCCTCAGAT 285
|||||
33 rGlyGlyGluCysTrpAlaPheAlaIleHisGlnAsnGlyIleuSerVal 49
|||||
286 CAACGGGGAATGCTGGCCATTTCCGCTCATCCAAAGACCGACGACAT 335
|||||
50 ValGlnAsnGlyMetTyrlsGlyPheThrLeuSerGluLeuTrpLeu 66
|||||
336 GTTGCAATAGCCCGCTATAAAGAAAGACATGTGACGCTTTGGGAAGA 385
|||||
66 sHisArgHisLeuPheGlyGlnLeuGluGlyAspArgPheProLeuLeu 83
|||||
386 GCACCGGTAAATATTCGCGCGGTAGAGGGGATCGGTTCCCGCTCTGA 435
|||||
83 hLysIleLeuAspAlaAspGlnAspLeuSerValGlnValHisProAsn 99
|||||
436 CAACCTGCTGATGTGAGGAATACGTCAATTAAGTTCAACCTGAT 485
|||||
100 AspGluTyrlaAsnIleHisGlnAsnGlyGluLeuGlyLysThrGlu 116
|||||
486 GTTTCCTATGCGGAGAAACGAGAGGAGAACTCGGCAAGAGGAATG 535
|||||
116 sTrpTyrlleIleAspCysGlnLysAspAlaGluIleIleTyrlGlyHis 133
|||||
536 CTGTACATATTACGCTGTAGGAAACGCAAAATCATTTAGGCGCAT 585
|||||
133 sPheThrThrLysGluGluLeuThrThrMetIleGluArgGlyGluTrp 149
|||||
586 CGGCGCGCTCAAAACGAACTGTCAATGATGACACACGCTGACTG 635
|||||
150 AspGluLeuLeuArgValValLysProGlyAspPhePheTyrla 166
|||||

```

```

636 GAGGCGCTGCGGAGACATCAAAATTAAACCGGTGATTTCATATATGT 685
166 1PserSerGlyThrValHisAlaIleGlySerGlyIleLeuAlaLeuGluT 183
686 GCCAGGCGGAACGCTGCACCATTTGTCAGAGGGGCCCTGTTTATGAGA 735
183 hrcGlnGlnAenSerAspThrThrTyrArgLeuTyrAspTyrAspArgLys 199
736 CTCGCAAAATTCAGATGCCATACCGGGTGTACATTTATGACCGCTCTT 785
200 AspaIaGluGlyLysLeuArgGluLeuHisLeuLysLysSerIleGluVa 216
786 GATGCAACGCAAGACTCCGAGAGCTTCATTTGCCAAGCGGTCAATGC 835
216 111Glu1Val1ProSerIleProGluArgHisThrValHisHisGluGlnI 233
836 CGCCACGGTTCCTCCATGTGCAGCGGTATATAGATGAAATCGACAGATCA 885
233 1eGluAspIleuLeuThrThrThrLeuIleGluGlySAlaTyrPheSerVal 249
886 GAAAGCAATACCATTAACATTTTGTCCAGGGCAATATTTTTCGTT 935
250 GlyTyrTrpAsnLeuSerGlySerAlaSerLeuLysGlnGlnLysProPh 266
936 TATATAATGGGACATCAATGCGGAGATGTAATGCTCAGATGAATCCTT 985
266 eleuLeuIleSerValIleGluGluGlyArgMetIleSerGlyGluT 283
986 TCTATTTTGGCGCTGATAGAGAGCGCTTGTCTCAATATAGAGACA 1035
283 yVal1TyrProPheLysLysGlyAspHisMetLeuLeuProTyrGlyLeu 299
1036 AAACATGTCGCCCTCAAAAAGGTGATCATTTATTTGGCGGCTCAAAAG 1085
300 GlyLysLysLeuLysGluGlyTyrAlaGluCysIleValSerHisLeu 315
1086 CCCCATTTACGATAAAGAACTGTACCTTATGTCTGTCTCATATT 1133

seq_name: gb_ba2:BACORFX
seq_documentation_block: 4162 bp DNA BCT 01-FEB-2000
LOCUS BACORFX
DEFINITION Bacillus subtilis gene for beta-N-acetylglucosaminidase, complete
cds.
ACCESSION D45048.1 GI:1129072
VERSION D45048.1 GI:1129072
KEYWORDS endo-beta-1,4-glucosaminidase; beta-N-acetylglucosaminidase.
SOURCE Bacillus subtilis (strain AC327) DNA, clones PEH4 and pMCEV4.
ORGANISM Bacillus subtilis
Bacteria; Firmicutes; Bacillus/Clostridium group; Bacillus.
REFERENCE
1 (bases 1 to 4162)
Margot, P., Muel, C. and Karamata, D.
The gene of the N-acetylglucosaminidase, a Bacillus subtilis 168
cell wall hydrolase not involved in vegetative cell autolysis
Mol. Microbiol. 12 (4), 535-545 (1994)
JOURNAL MEDLINE
REFERENCE
2 (sites)
Rashid, M.H., Mori, M. and Sekiguchi, J.
Glucosaminidase of Bacillus subtilis: cloning, regulation, primary
structure and biochemical characterization
Microbiology 141 (Pt 10), 2391-2404 (1995)
JOURNAL MEDLINE
REFERENCE
3 (bases 1 to 4162)
Sekiguchi, J.
Direct Submision
Submitted (07-JAN-1995) to the DDBJ/EMBL/GenBank databases.
JOURNAL MEDLINE
REFERENCE
4 (bases 1 to 4162)
Sekiguchi, J.
Department of Applied Biology, 3-15-1 Tokida, Ueda-shi, Nagano
Japan (E-mail: jsekiguchi@ipc.shinshu-u.ac.jp, Tel: 0268-21-5344,
Fax: 0268-21-5331)
FEATURES
source
1. 4162
Location/Qualifiers

```

```

-35_signal
-10_signal
RBS
gene
CDS
stem_loop
-35_signal
-35_signal
-10_signal
-10_signal
-10_signal
RBS
sig_peptide
gene
CDS
/organism="Bacillus subtilis"
/strain="AC327"
/db_xref="taxon:1423"
/clone="PEH4 and pMCEV4"
34..39
/note="Sigma-A type"
61..66
/note="Sigma-A type"
200..208
/gene="OrfX"
217..1167
/gene="OrfX"
217..1167
/gene="OrfX"
/codon_start=1
/transl_table=11
/product="unknown"
/protein_id="BAA08088.1"
/db_xref="GI:1129073"
/translation="MTSPFIFLPVPEKRIWGGTALRDPRGVSIPSESTGECMAISAH
PKGSTVANGCYKKGKTLIPMEHREVRVGVGVEDRDELPLKLDVREDSIVHPDY
VAGNEREBELKTCWYIIDCKENAEIYHTARSTELIYMTINSGDEGLIRIKYK
PGRHYVPSGLHALCKGALYLETOQNSDARTYIDIDRLDSGSRREIHPAKAYMA
TVPHYDGTIDESTSRKGTITKTFVQCEIVSVTRWMDINGAEAAODESFLICVLDGS
GLKXEDTCTCLKGDHPTLPQMPDETITCTLTYSRL"
1166..1194
/note="rho-independent terminator"
/evidence=not_experimental
1224..1228
/note="sigma-D type"
1241..1246
/note="sigma-A type"
1244..1250
/note="sigma-D type"
1265..1270
/note="sigma-A type"
1285..1291
1296..1376
/gene="CwlG"
/product="27 amino acids long signal peptide"
1296..3938
/gene="CwlG"
1296..3938
/gene="CwlG"
/EC_number="3.2.1.30"
/function="hydrolyze beta-1,4 glycan bond between the
GlcNAc & MurNAc"
/codon_start=1
/transl_table=11
/evidence=experimental
/product="beta-N-acetylglucosaminidase"
/protein_id="BAA08089.1"
/db_xref="GI:1129074"
/translation="MKKRILAPMLLSAASLAFPMSSGSAQAAYTDYSIKYKPSNFT
STESASQAVAKLENDTGMDASTYDASGITTTTQTSAGHISEBAKAIISGLAKQSTI
TGTSVPGSSKPYTITSGATISGKQANTILAKIQTGVAGAKATGAQPEPMWMT
SDIADERTVKALIQSLAKQTGKSSYQPIHTYVSVTTQSTIVGSGRAQIKNAPQK
ESGLOASLKEITVKGQAVYETTTAAIISBANAKTLLDQLKQSTGITSYSINQKTVE
SYNVOASAFKGLSTVKDAISOIKKNTVSGSYQVGSYTNMGGITKQDQKIDT
FEKKKKHWYTSVSKKTTTSAAYQITTAKLIGEQANKAAAPFAQKRYAKAAGST
AENQQLSEETSOQAVTKGILNKLKQLSASAKSVKQIADTFKITTESLIDQKY
NOALTFESNHSIVASOKTGOTAAASYQITTEALISQEDRIDVLFKKNHVSITK
TGQATYQYKIVTTQLSKRTALNGLNGLSKSVTPETYSNTLTKISVNEQFTND
TAAASRTKLQLYGTMTSIVYIKKQPMKNTNLSLRDVOKQMTVSPOTGAAVYS
LTVINATSTYATAVINIRSPREYSPNVIQGRKDKPKVYAGCGIITGPKAFID
AANQYSINELIISHALLETNGTSLANKLITNGKTYVMYNTGICAYDSNPYQKYK
ATQDQWTFPEAAITGGAFTQSSSTINATATNODTLKRMKSATATHQIATIDGMAIKQ
VNRMSLSDLDGTYTLFDVPEYK"
1377..3935
/gene="CwlG"
/EC_number="3.2.1.30"
/note="853 amino acids long mature peptide"
/function="hydrolyzes beta-1,4 glycan bond between GlcNAc

```


TITLE Direct Submission
JOURNAL Submitted (18-NOV-1997) I. Moszer, A. Danchin, Institut Pasteur,
Regulation de l'Expression Genetique, 28 rue du Docteur Roux, 75724
Paris Cedex 15, FRANCE. E-mail: moszer@pasteur.fr,
adanchin@pasteur.fr Phone: +33 (0)1 45 68 84 41, Fax: +33 (0)1 45
68 89 48

FEATURES
source Location/Qualifiers

1. 212610
/organism="Bacillus subtilis"
/strain="168"
/db_xref="taxon:1423"
complement(10. .3879)
/gene="yvnB"
complement(10. .3879)
/gene="yvnB"
/function="unknown"
/codon_start=1
/transl_table=11
/protein_id="CAB15521.1"
/db_xref="GI:2636030"
/db_xref="SPTREMBL:O34986"
/translation="MRKYTVASILLISFLSYLSGHHESKAFPYVOGELRQTPHYPL
LENPROIDVKAEMSNLMTEAOVRDNSYOAFAFLFKOSNELGYKRVVPDPTGAVR
KFIQIPKRIINSELEYITMSNGKORVSEETKILLEGQADLAHIFELLITELV
DTKNIGRADGYEFIEFTTTRTDYDFKHYHRYRYKKEGOSDLIMRDEKRLVSGE
TFVVALRAPHGELETSADPNRTYOTQLKEGKNLAVIDETBOMANTPRRAVISTNTEK
DISVAHKKHALRLSSVLKIPPLNGLIELNLISIGERNPSYRTDQAOVDDKRIK
LDKEKPVLEDLDRKVPRAVSELRLADIRDSRLVKTAVAFYRTDENKPKRLAEKD
RNDLSEGTFLKATTEGRADSIKMLIDGRKQTEPAMEKEVYFAFDRKRLYKRNAT
MEGKLVAFDDTNTKRYTSVLPETLLRKQKQIYIRSGSVSPEDFAENADDFL
VKNALVLSGDTVIRDOVRSEKELFTGDNORSNKSMQFNLPSGLFTSQLEMDIS
KLESGAHIQASDGENSVLVYVDSNGPHIEPNITGQYKGNLILQAMKRWKMSIT
EEASGLGESITLLPYHTSSDLLPGKHSLLKVTATDLGKNKTVIERLEKFERHSDP
EVIDEADNTHAKLSVRYKQDPTNDAMPDGFGRQYFARHNVKIFKIASLTPKPSY
PERETPTNKELETERYSADGKTVSTENKELPPYRPFVTDPSSIDENDLAEVWSS
LPGRKVTWYANNYRTNEMQPYDSFAKDDKPFILKASVIADPVRESKRVYIDEP
PAKMDYVWMSDPOYVAESYPIEDKQTEKIKRNOKQIMIKYVPHGDIYDSDADR
OKNADBSMSYLDGTPYGLAENHVGKDSYRFGKTRGSDRDKDFPHGGSK
NNRGHYDITSSNGNDYIMLNGKQITBEDLAKMQVLRKHPDRAIIAFHEYLVSQ
NSPIGEKIFKEIVKPNVAVVLSGHTHSARKTDELDGDKPKRLVQHMADYOG
GPEGQGYLRLLQFQANDMVHVSISPYVDRKMYVDTYVGNKDESSLSLDKPRK
KVEIDYFECNVYITNEELGRBQVKSQDAEFRWDLBPQSVYVYVILVEDSPFGKRS
PIWKEKIKETRYRPAPODFDRHVSNP"
4044. .4517
/gene="yvna"
4044. .4517
/gene="yvna"
/function="unknown"
/note="similar to hypothetical proteins from B. subtilis"
/codon_start=1
/transl_table=11
/protein_id="CAB15522.1"
/db_xref="GI:2636031"
/db_xref="SPTREMBL:O34692"
/translation="MDSIHDELFOAIOOFAIKKDKRVKQVQPSIGISSQHHDLIK
KDWLTQIAIVSCHIHTSONVNSFLASRLHSKAASVKAVALKHNIIIVTKRPGK
KEIFTYLTDSGRKLAALHEQJHEKAKKQYKOLFNEFSIDDLKVTATFAFNIMIKYM"
complement(4558. .5775)
/gene="cydX"
complement(4558. .5775)
/gene="cydX"
/codon_start=1
/transl_table=11
/product="cytochrome P450-like enzyme"
/protein_id="CAB15523.1"
/db_xref="GI:2636032"
/db_xref="SPTREMBL:O34926"
/translation="MSOSIKLFSVLISQFOFNNPYAFPSQLREDDPVHEESIDSYFIS
RYHDVRYLQHPDILFTFKSLERAPVNRGVSVAQMGKESAKRRIVVFSFKGALD
HLSPLIKONAHNPDLFTFKSLERAPVNRGVSVAQMGKESAKRRIVVFSFKGALD
VADFTTISQSPERASIMSEQLSOLIMPVIKERVNPNGSDLSITCTSEVEGAL
SDKDILALILNVLATERPDKTLAMIMYHLNNPQMDVNLADRSIVRAIMETLRY

gene
CDS

gene
CDS

gene
CDS

terminator

gene
CDS

gene
CDS

KPPVOLIPROLSQDTVAGMEIKDQITVECMIGANRDPFAEPDVFNIHREDJGK
SAFSGAARHLAFSGIHNCGVGAFAKNEIEIVANIVLDKMRNIRLEEDFCYAESGLYT
RCPVLLVAFDA"
complement(5791. .6537)
/gene="yvmC"
complement(5791. .6537)
/gene="yvmC"
/function="unknown"
/note="similar to hypothetical proteins"
/codon_start=1
/transl_table=11
/protein_id="CAB15524.1"
/db_xref="GI:2636033"
/db_xref="SPTREMBL:O34351"
/translation="MTGMVTERRSYVHFAEALTFENCREIEPERRRHLVGLISPNRPS
EDYIYRLIGMAKQFVSVLLAGHEANLLEALGTPRGKAERKVRKESRNRFAAR
ALVAGDPRKAIHFPSPFDINKKAYOLRQVEHAFEPQPHRHACLDMSREALIGRAR
GVSLMEVESSEDMNLAVEYVIAELPFPIGAPDILEVEETLLAVHRPMKGEKISNHE
FSICKRPNQGYIVQEMQMSSEKRTSEG"
6963. .7472
/gene="yvmB"
6963. .7472
/gene="yvmB"
/function="unknown"
/note="alternate gene name: yzha
similar to hypothetical proteins"
/codon_start=1
/transl_table=11
/protein_id="CAB15525.1"
/db_xref="GI:2636034"
/db_xref="SWISS-PROT:P40762"
/translation="MSDLTKOMIYDIYRLHLNEOKANSLQOOFKEAAEDVAEIP
KMTSIIHYDCGHEPFINNNGIARKMLSKANYTKISTRIKEEPIINSYOLDNKE
VYFKLRRGQYFDFLHEHLKKELAFYQGLFSDPSQDEQKAVLKEFLQSLSTLEAGT
DSTPRPKPK"
7493. .8726
/gene="yvna"
7493. .8704
/gene="yvna"
/function="unknown"
/note="similar to transporter"
/codon_start=1
/transl_table=11
/protein_id="CAB15526.1"
/db_xref="SPTREMBL:O34307"
/translation="MAHTKSKAVLILVYVCSAFASISQNIYSPIILIKESPHYST
AMVNLVSVEFMVITPAIMQIITIGAITDEGARIYITGIIANASISGCAVTDPTFLT
IFRMIOAGSAPILPILATTTGOLFTGERSAMGTOMILSVAPAIAPVIGFIGA
ACVGEIPEMLAISIVLYTNSITFPKDSPTESQOKKGVFAHRSKISFTNRGNYIL
TISPLFTFIAVYIYPLTLEHRTIDVAGLILPLASTIAGTFLKRIQAKIG
LHTPLFGSNVIAACSIILFAVTHSVSLVMAVLTALFGISGVTIPPLVSTWITNEBH
NKGSAIGMNFRTRYGMAAPVSAVLLTTPMSMPSLLGLGFAALSCFLLPMSFP
KRTKQKRNHM"
8709. .8726
/gene="yvna"
complement(8732. .9091)
/gene="yvID"
complement(8732. .9091)
/gene="yvID"
/function="unknown"
/codon_start=1
/transl_table=11
/protein_id="CAB15527.1"
/db_xref="GI:2636036"
/db_xref="SPTREMBL:O34648"
/translation="YKNAVASILVNALLIVIDGYIDSIRHISGAIINSLITILN
VFLIKLITFLPVTMYVLGLFLVINAITLMTASIMGDSFOIDGCGTAIMASVILLS
VFLIKLIGSIEPLRKK"
complement(9093. .9290)
/gene="yvlC"
complement(9093. .9290)
/gene="yvlC"

```

/function="unknown"
/codon_start=1
/transl_table=11
/protein_id="CAB15528.1"
/db_xref="GI:2636037"
/db_xref="SPRMBL:034719"
/translation="MNKLYRSEKNKTIAGVIGLAEYFNMDASLLRVTIVLLAINTSV

```

```

alignment_scores:
  Quality: 978.50      Length: 316
  Ratio: 3.822        Gaps: 1
  Percent Similarity: 81.013  Percent Identity: 56.646

```

alignment_block:

US-09-462-846-2 x BSUB0019/rev ..

Align seg 1/1 to reverse of: BSUB0019 from: 1 to: 212610

```

1 MetThrThrGluProLeuPhePhePhePheProValPheIleGluArgIleT 17
||||| :|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
90517 ATGCGCAATACCGCATTTTCTACGCCCTGTGTTAAAGAAAATCTG 90468
17 pgiLyGlyThrAlaLeuAlaSP...PheGlyTYrThrIleProSerGlnA 33
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
90467 GGGCGGGAACCGCTTTACGAGATGATTGGATACAGATTCCTTCAGAA 90418
33 rGThrGlyGluCysTrpAlaPheAlaAlaHisGlnHisGlnGlyInserVal 49
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
90417 CAACGGGGGAATGCTGGCCATTCTCCGTCATCCAAAGACCGAGCACT 90368
50 ValGlnAsnGlyMetTYrLYrGlyPheThrLeuSerGluLeuTrpGluH 66
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
90367 GTTGCAAAATGCCCGGTATTAAGAAAACATTGATCAGCTTTGGGAGA 90318
66 SHLSArgHisLeuPheGlyGluLeuGlyAspArgPheProLeuLeuT 83
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
90317 GCACCGGGAATGATTCGGCGGTAGAGGGGATCGGTTTCCCTCTGCA 90268
83 hTrpSileuAspAlaSPGlnAsnLeuSerValGlnValHisProAsn 99
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
90267 CAACGCGCGGATGTGAAGAGATACGTCATTAAGTTTCAACCCGAT 90218
100 AspGluTYrAlaAsnIleHisGluAsnGlyGluLeuGlyTYrGlyC 116
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
90217 GATTACATGCGCGGAAAGAGAGAGGAACTCGGCAAGACGGAAATG 90168
116 STPrTYrIleLeuAspCysGlnAspAlaGluIleIleTYrGlyHisA 133
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
90167 CTGTACATATATCGATGTAAAGAAAACGCAAAATCATTTACGGCATA 90118
133 snAlaThrTrpThrGluGluLeuThrThrMetIleGluArgGlyGluTr 149
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
90117 CGGCGCCGCTCAAAAACGAACTTGTCAATGATCAACACGCGTACG 90068
150 AspIleuLeuAlaArgValIleValysProGlyAspPhePheTYrVA 166
::: |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
90067 GAGGCGCTGCGCGAAGATCAAAATTAACCGGATATTTCTATTTATGT 90018
166 IProSerGlyThrValHisAlaIleGlyGlyGlyIleuAlaLeuGluT 183
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
90017 GCGGAGGGGAACGCTGCACGATGTGCAAGGGGCCCTGTTTATAGAGA 89968
183 hTrGlnGlnAsnSerAspThrThrTYrArgLeuTYrAspTYrAspArgLY 199
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
89967 CTGACCAAAATTCAGATGCCACATACCGGGGTACGATTAAGACCGCTT 89918
200 AspAlaGluGlyLYrLeuArgGluLeuHisLeuLYrSerIleGluVA 216
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
89917 GATAGCAACGGAAGTCCGAGACCTCATTTTGCAGGCGGTCATATGC 89868
216 ILeGluValProSerIleProGluArgHisThrValHisHisGluIn 233
::: |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
89867 CGCCACGGTTCCTCCATGTGACGGGTATATGATGAAATCGACAGATCA 89818

```

```

223 LeuIleuPheLeuThrThrThrIleuIleGluCysAlaTYrPheSerVal 249
::: |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
89817 GAAAGGATATACCATTAACATTTGTCCAGGGGATATTTTCGGTT 89768
250 GlyLYrTrpAsnLeuSerGlySerAlaSerLeuLYrGlnGlnLYrProPh 266
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
89767 TATAATGAGCATCAATGCGAACCTGAAATGGCTCGAGGATGAATCCT 89718
266 eLeuLeuIleSerValIleGluGlyGluArgMetIleSerGlyIuT 283
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
89717 TCTGATTTGCGACGCGATGAGAGGAGCGGTTGCTCAAGTATGAGACA 89668
283 yTrValTYrProPheLYrGlyAspHisMetLeuLeuProTYrGlyLeu 299
::: |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
89667 AAACATGCTCCGCTCAAAAAGGATGATCACTTATTTTGGCGGCTCAATG 89618
300 GlyGluPheLYrLeuGluGlyTYrAlaGluCysIleValSerHisLeu 315
::|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
89617 CCCGATTTACGATTAAGAACTGTACCTTATCGTGTCTCATATTT 89570
seq_name: gb_pat1:AX006205

```

```

seq_documentation_block:
LOCUS AX006205 945 bp DNA PAT 24-AUG-2000
DEFINITION Sequence 6 from Patent WO9904016.
ACCESSION AX006205
VERSION AX006205.1 GI:9929073
KEYWORDS
SOURCE
ORGANISM
.
Bacillus subtilis.
Bacillus subtilis
Bacteria: Firmicutes; Bacillus/Clostridium group;
Bacillus/Staphylococcus group; Bacillus.
REFERENCE
1 (bases 1 to 945)
AUTHORS
Estell,D.A.
TITLE
Proteases from gram-positive organisms
JOURNAL
Patent: WO 9904016-A 6 28-JAN-1999;
GENECOR INTERNATIONAL B V (NL); ESTELL DAVID A (US)
FEATURES
source
1..945
/organism="Bacillus subtilis"
/db_xref="taxon:1423"
BASE COUNT 290 a 189 c 223 g 243 t
ORIGIN

```

```

alignment_scores:
  Quality: 958.50      Length: 313
  Ratio: 3.804        Gaps: 1
  Percent Similarity: 80.511  Percent Identity: 55.591

```

alignment_block:

US-09-462-846-2 x AX006205 ..

Align seg 1/1 to: AX006205 from: 1 to: 945

```

3 ThrGluProLeuPhePheLYrProValPheLYrGluArgIleTrpGly 19
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
4 ACCATCCATATATTTAGAGCCCTGCTTAAAGAAAGCTATAGGGGAGG 53
19 yTrpAlaLeuAlaSP...PheGlyTYrThrIleProSerGlnArgThrG 35
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
54 GACGAGCTTCGAGCGCTTTGGCTAGCGAATACCTCGCAAAAACAG 103
35 LYrGlyCysTrpAlaPheAlaAlaHisGlnAsnGlyGlnSerValIleG 51
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
104 GTAAGTCTGGCGCTTCTGCACATGCCCATGCTGCTGCTGTGAAGA 153
52 AsnGlyMetTYrLYrGlyPheThrLeuSerGluLeuTrpGluHisIse 68
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
154 AATGGCCCGCTGGCAGGAAAGACATGATGATCAAGTATGCAAGATCAT 203
68 gHisLeuPheGlyGluLeuGlyLYrAspArgPheProLeuLeuThrLY 85

```



```

85  leleaspaalaspGlnAspLeuSerValGlnValHisProAsnAspGlu 101
    ::::::::::::::::::::::::::::::::::::::::::::::::::::
18079 TGGCGAGCGCCAAATGATGATCTCTCCGTCGAACTCCATCTCTGATGATGAT 18128
102  TyrAlaAsnIleHisGluAsnGlyGluLeuGlyLysThrGluCysTrpTyr 118
    ::::::::::::::::::::::::::::::::::::::::::::::::::::::
18129 TATGCAGAAATGCTGCAGAAATGGCAGCTTGTATTAACGGAGTGTGGTA 18178
118  rlieilaspcysgslmlyaspAlaGluIleIleTyrGlyHisAsnAlaT 135
    ::::::::::::::::::::::::::::::::::::::::::::::::::::::
18179 TATCATTTGATTCGAAAGATGACCGCCAGACTTAATTTGGGACATACAGCA 18228
135  hrThrLysGluGluLeuThrThrMetIleLysArgGlyLysTrpAspGlu 151
    ::::::::::::::::::::::::::::::::::::::::::::::::::::::
18229 GCACAAAGAGAGAGTTCAACCAACGATTAACAAAGCGTGTGGAAACGGC 18278
152  LeuLeuArgArgValLysValLysProGlyAspPhePheTyrValProse 168
    ::::::::::::::::::::::::::::::::::::::::::::::::::::::
18279 CTGCTGAGCGCAATCAAAATCAAGCCGAGATTTCTTTATGTGCGCAG 18328
168  rglYThrValHisAlaIleGlyLysGlyIleLeuAlaLeuGluThrGlnG 185
    ::::::::::::::::::::::::::::::::::::::::::::::::::::::
18329 CGGACACCTCATGCTTATGCTTAAGGACCCCTTCTCTTGAATCCAGC 18378
185  lmsAsnSerAspThrThrThrArgLeuTyrAspTyrAspArgLysAspAla 201
    ::::::::::::::::::::::::::::::::::::::::::::::::::::::
18379 AAAACTCTGATACAAACATATCGGCTATACGATTAATGACCGCTGTAATGAC 18428
202  GluGlyLysLeuArgGluLeuHisLeuLysLysSerIleGluValIleG 218
    ::::::::::::::::::::::::::::::::::::::::::::::::::::::
18429 CAGGCGCAAAAGAACTCTTCATATGAGAAAGACCATGGAATGATCAATAC 18478
218  vvalProSerIleProGluArgHisThrValHisHisGluGlnIleGluA 235
    ::::::::::::::::::::::::::::::::::::::::::::::::::::::
18479 GATACCGCATATGATCAATAGCTATACACCGGAGTAAGTAAGAAAGTTGGTA 18528
235  spledLeuThrThrThrLeuIleGluCysAlaTyrPheSerValGlyLys 251
    ::::::::::::::::::::::::::::::::::::::::::::::::::::::
18529 ACGCTGATGATCATTTGTTATGTCATACGATTAATTTCTCAGGTGTAACA 18578
252  TrpAsnLeuSerGlySerAlaSerLeuLysGlnGluLysProPheLeu 268
    ::::::::::::::::::::::::::::::::::::::::::::::::::::::
18579 TCGAAGATTTAGCGCGCCGACCTCTTCTTCATATCAAAACATATTGCT 18628
268  uileSerValIleGluGlyLysGluArgMetIleSerGlyLysValT 285
    ::::::::::::::::::::::::::::::::::::::::::::::::::::::
18629 GGGGAGTCTCTGACGCGCATCAGACGAATCAATTAATATGATTCAGT 18678
285  ytrProPheLysGlyAspHisMetLeuLeuProTyrGlyLeuGlyGlu 301
    ::::::::::::::::::::::::::::::::::::::::::::::::::::::
18679 ATGAAATCAATGCAAGGCTCACACTTATTTCTGCTCGCCGCAATTTTGGAGA 18728
302  PheLysLeuGluGlyTyrAlaGluCysIleValSerHis 314
    ::::::::::::::::::::::::::::::::::::::::::::::::::::::
18729 TTTTACATAGAGAAACATGTAATTCATGATATCTCAT 18767

seq_name: gb_daz:BSUB0004
seq_documentation_block:
LOCUS      BSUB0004      213190 bp      DNA      BCF      26-NOV-1997
DEFINITION  Bacillus subtilis complete genome (section 4 of 21): from 600701 to
            813890
ACCESSION  Z99107 AL009126
VERSION    Z99107.1 GI:2632866
KEYWORDS
SOURCE
ORGANISM   Bacillus subtilis.
            Bacteria; Firmicutes; Bacillus/Clostridium group;
            Bacillus/Staphylococcus group; Bacillus.
REFERENCE  Kunst, F., Ogasawara, N., Moszer, I., Albertini, A.M., Alloni, G.,
            Azevedo, V., Berrtero, M.G., Bessieres, P., Bolotin, A., Borchert, S.,

```

```

Boriss, R., Boursier, L., Brans, A., Braun, M., Bignell, S.C.,
Bron, S., Brouillet, S., Brusch, C.V., Caldwell, B., Capuano, V.,
Carter, N.M., Choi, S.K., Codani, J.J., Conerton, I.F., Cummings, N.J.,
Daniel, R.A., Denzlot, F., Devine, K.M., Dusterhoft, A., Ehrlich, S.D.,
Emmerson, P.T., Enlian, K.D., Errington, J., Fabret, C., Ferrati, E.,
Fouger, D., Fritz, C., Fujita, M., Fujita, Y., Funa, S., Galizzi, A.,
Galleron, N., Ghim, S.Y., Glaser, P., Goffeau, A., Goldlighty, E.J.,
Grandi, G., Guiseppi, G., Guy, B.J., Haga, K., Halech, J., Harwood, C.R.,
Henaut, A., Hilbert, H., Holsappel, S., Hosono, S., Hullo, M.F.,
Itaya, M., Jones, L., Joris, B., Katamata, D., Kaishata, T.,
Klaerr, Blanchard, M., Klein, C., Kobayashi, T., Koetter, P.,
Kontagstein, G., Krogh, S., Kumano, M., Kurita, K., Lapius, A.,
Lardinois, S., Lauber, J., Lazarevic, V., Lee, S.M., Levine, A., Liu, H.,
Masuda, S., Maue, C., Medigue, C., Medina, N., Noone, D., O'Reilly, M.,
Ogawa, K., Ogiwara, A., Oudega, B., Park, S.H., Parro, V., Pohl, T.M.,
Portetelle, D., Portwollik, S., Prescott, A.M., Prescan, E., Pujic, P.,
Purnelle, B., Rapoport, G., Rey, M., Reynolds, S., Rieser, M.,
Rivolta, C., Rocha, E., Roche, B., Rose, M., Sadate, Y., Sato, T.,
Scanlan, E., Schleich, S., Schroeter, R., Scoffone, F., Sekiguchi, J.,
Sekowska, A., Seror, S.J., Serron, P., Shin, B.S., Solido, B.,
Sorokin, A., Tacconi, E., Takagi, T., Takahashi, H., Takemaru, K.,
Takeuchi, M., Yamakoshi, A., Yanaka, T., Terpsstra, P., Tognoni, A.,
Tosato, V., Uchiyama, S., Vandenbol, M., Vannier, F., Vassaret, A.,
Viari, A., Wambutt, R., Wedler, E., Wedler, H., Welleneger, T.,
Winters, P., Wipat, A., Yamamoto, H., Yamane, K., Yasumoto, K., Yata, K.,
Yoshida, K., Yoshikawa, H.F., Zumbstein, E., Yoshikawa, H. and
Danchin, A.

The complete genome sequence of the gram-positive bacterium
Bacillus subtilis
Nature 390 (6657), 249-256 (1997)
98044033
2 (bases 1 to 213190)
Kunst, F., Ogasawara, N., Yoshikawa, H. and Danchin, A.
Direct Submission
Submitted (18-NOV-1997) I. Moszer, A. Danchin, Institut Pasteur,
Regulation de l'Expression Genetique, 28 rue du Docteur Roux, 75724
Paris Cedex 15, FRANCE. E-mail: moszer@pasteur.fr,
adanchin@pasteur.fr Phone: +33 (0)1 45 68 84 41, Fax: +33 (0)1 45
68 89 48

FEATURES
            source
            Location/Qualifiers
                1..213190
                /organism="Bacillus subtilis"
                /strain="168"
                /db_xref="taxon:1423"
                33..740
                /gene="ydf5"
                33..740
                /gene="ydf5"
                /function="unknown"
                /note="similar to hypothetical proteins"
                /codon_start=1
                /transl_table=11
                /protein_id="CAB12373.1"
                /db_xref="GI:2632867"
                /db_xref="SPTREMBL:P96697"
                /translation="MIELEVYIRIVASFGILLIERIKKQIISOMTIFDIATITG
                ARAAGATNTSIKPHNMAISFIVLITFLIFLISPLSKRKRFKADPVLIONGI
                LESNMKRYTLIDYLNQOLREKEIPEVEVFAITEENGLQTLVRKQFRIYKODIM
                IAVNQRPIELIMDGEIENLNKQNLFTSWLIEELRKRDINVKETVAVILGND
                IYVDYKHQISYPMKE"
                complement(755..1186)
                /gene="ydf1"
                complement(755..1186)
                /gene="ydf1"
                /function="unknown"
                /codon_start=1
                /transl_table=11
                /protein_id="CAB12374.1"
                /db_xref="GI:2632868"
                /db_xref="SPTREMBL:P96698"
                /translation="MDPEKIKRMLEITITRYOSDFWTVLYKAKPEHFDSEASTPY
                DTQDEINFTIVEMGYIEELITRLSKTQLLIKGITITPFAEMSEVLERIKETIGEL

```



```

US-09-462-846-2 x AE006310 ..
Align seg 1/1 to: AE006310 from: 1 to: 10607

4 GluProPhePhePhePheProValPheLysGluArgIleTyrPglYglYnh 20
|||||.....:|||||.....:|||||.....:|||||.....:
7717 GAACCATGTTGTTGACSTGATTTGGACAGAAAATAATCGGGCGCGCA 7766

20 rAlaLeuAlaAspPheGlyTyrThrIleProSerGlnArgThrGlyGluC 37
|||||.....:|||||.....:|||||.....:|||||.....:
7767 CCATTGTAAGAGTTGGCTATGATTTGGCATGACAAAGTTGGTGAAT 7816

37 ysrPAlaPheAlaAlaIAsnGlnsnglyInserValValGlnsngly 53
|||||.....:|||||.....:|||||.....:|||||.....:
7817 ATGGGCTATTTGCTGATCCATCCATGCTGTGCAACAATTCCTAATGGC 7866

54 MetTyrLysGlyPheThrLeuSerGlnLeuTyrPglYnhIshArgHisE 70
|||||.....:|||||.....:|||||.....:|||||.....:
7867 GAATTTAAAGTCAAAAAGTTGACCATTTATGACGAAGTCACCGCCAAAT 7916

70 uPheGlyGlnLeuGluGlyAspArgPheProLeuLeuThrLysIleLeuA 87
|||||.....:|||||.....:|||||.....:|||||.....:
7917 GTTGGTGATAGTAAAAAGAGTTTTCCTTACTAATAAATTTTATAG 7966

87 sPAlaAspGlnAspLeuSerValGlnValHisProAsnAspGluTyrAla 103
|||||.....:|||||.....:|||||.....:|||||.....:
7967 ATGCCAATGACTGGCTTTCTGTGCAAGTTCATCCAGATGAATATGGA 8016

104 AsnIleHisGlnsnglyGluLeuGlyLysThrGluCysTyrTrpIleIle 120
|||||.....:|||||.....:|||||.....:|||||.....:
8017 CAAAACATGAA...GGTGAACCTGGAAAAACATGATGTTGGTACATAT 8063

120 eAspCysGlnLysAspAlaGluIleIleTyrGlyHisAsnAlaThrL 137
|||||.....:|||||.....:|||||.....:|||||.....:
8064 TTCAGCTGAACAGCGCTGAGATATCTATGACAAAGCAATTAATCC 8113

137 ysgLgluLeuThrThrMetIleGluArgGlyLysTyrAspGluLeuLeu 153
|||||.....:|||||.....:|||||.....:|||||.....:
8114 GTCAAGATTTAGCAGAAATGATTAATCTGTGATGGATCATTTGTGA 8163

154 ArgAlaGlyValLysValLysProGlyAspPhePheTyrValProSerGly 170
|||||.....:|||||.....:|||||.....:|||||.....:
8164 CGTAGGGTAAGTGAACAGGGGATTTCTCCATGTCTCGTCAGAGAAC 8213

170 rValHisAlaIleGlyLysGlyIleLeuAlaLeuGluThrGlnGlnAsn 187
|||||.....:|||||.....:|||||.....:|||||.....:
8214 AATGACGCAATCGGTGCTGGAATTTGTTATTTCTTGAACACAAATCT 8263

187 eAspThrThrTyrArgLeuTyrAspTyrAspArgLysAspAlaGluGly 203
|||||.....:|||||.....:|||||.....:|||||.....:
8264 CTGATACAACTTACCGTGTATTATGATTTGCACCTAAATAATGACCAAGA 8313

204 LysLeuAlaGlyLeuHisLeuLysSerIleGluValIleGluValPr 220
|||||.....:|||||.....:|||||.....:|||||.....:
8314 AATCTAGCTGATTTACATATTCACAAACATTTGATATGTAATGTAATCC 8363

220 o.....SerIleProGluArgHisThrValHisGluGlnIleGluA 235
|||||.....:|||||.....:|||||.....:|||||.....:
8364 GGGCACAAGATCCCTGAA...AATCAATTTAAACCTGAAATTTCTGCTG 8410

235 sPLeuLeuThrThrThrLeuIleGluCysAlaTyrPheSerValGlyLys 251
|||||.....:|||||.....:|||||.....:|||||.....:
8411 ATGCAAAATTTACAACTTGTGAAATCAATTTCTTGTGATGTTTATAAA 8460

252 TyrAsnLeuSerGlySerAlaSerLeuLysGlnIleLysProPheLeu 268
|||||.....:|||||.....:|||||.....:|||||.....:
8461 TGGCAAAATTCATGTCGACCATGAAATTTACAAAGTTCGATTAACCTTT 8510

268 uIleSerValIleGluGlyGluGlyArgMetIleSerGlyGluTyrValT 285
|||||.....:|||||.....:|||||.....:|||||.....:
8511 AGTTCTGTACTTATGTCGTAAGGAAATTAACAATGATGAAATGAAT 8560

285 yPrProPheLysGlyAspHisMetLeuProTyrGlyLeuGlyGlu 301

```

```

|||||.....:|||||.....:|||||.....:|||||.....:
8561 ATCCAGTTGAAAAAGAGCTCATTTATCTTACCAAGACACATTTGAAAAA 8610

302 PheLysLeuGluGlyTyrAlaGluCysIleValSerHis 314
|||||.....:|||||.....:|||||.....:|||||.....:
8611 TGAATTTGCTGCTCAATTAAGAAATTAATTCACGCAAT 8649

seq_name: gb_htg25:SPNEU1904

seq_documentation_block:
LOCUS SPNEU1904 147006 bp DNA HTG 16-DEC-2000
DEFINITION Streptococcus pneumoniae clone G54, *** SEQUENCING IN PROGRESS ***,
in ordered pieces.
ACCESSION AL449926
VERSION AL449926.1 GI:11545151
KEYWORDS HTG; HTGS_PHASE2.
SOURCE Streptococcus pneumoniae.
ORGANISM Bacteria; Firmicutes; Bacillus/Clostridium group; Streptococcaceae;
Streptococcus.
REFERENCE 1 (bases 1 to 147006)
AUTHORS Dopazo,J., Mendoza,A., Herrero,J., Caldera,F., Polissi,A.,
Humero,J., Friedl,L., Guerrier,M., Grand-Schenk,E., Gandin,C., de
Francesco,M., Buell,G., Feger,G., Garcia,E., Peltich,M. and
Garcia-Bustos,J.F.
Direct Submission
Submitted (31-OCT-2000) Research Department, Glaxo Wellcome, S.A.,
Severo Ochoa 2, 28760 Tres Cantos, SPAIN
* NOTE: This is a 'working draft' sequence.
* This sequence will be replaced
* by the finished sequence as soon as it is available and
* the accession number will be preserved.
LOCATION/Qualifiers
Source 1..147006
/organism="Streptococcus pneumoniae"
/serotype="19F"
/db_xref="taxon:1313"
/clone="G54"

BASE COUNT 44762 a 26837 c 32084 g 43281 t 42 others
ORIGIN

alignment_scores:
Quality: 839.00 Length: 316
Ratio: 3.467 Gaps: 5
Percent Similarity: 76.582 Percent Identity: 53.797

alignment_block:
US-09-462-846-2 x SPNEU1904 ..
Align seg 1/1 to: SPNEU1904 from: 1 to: 147006

3 ThrGluProLeuPhePhePheProValPheLysGluArgIleTyrPglYgl 19
|||||.....:|||||.....:|||||.....:|||||.....:
79069 TCAGAACCATTAATTTTACATCTGATGACAAAGAAAATCTGGGGTGG 79118

19 yThrAlaLeuAlaAsp...PheGlyTyrThrIleProSerGlnArgThrG 35
|||||.....:|||||.....:|||||.....:|||||.....:
79119 AGCCAAAGCTAGCTAGTATGATTTGGCTTACGACATCCCAAGTGAATAATG 79168

35 IyGLyCysTyrPAlaPheAlaAlaIAsnGlnsnglyGlnSerValAlaGln 51
|||||.....:|||||.....:|||||.....:|||||.....:
79169 GAGATATTTGGGCGCATCTCACCCATCCAAATGAGTCTTAAGTTGCC 79218

52 AsnGlyMetTyrLysGlyPheThrLeuSerGlnLeuTyrPglYnhIshArg 68
|||||.....:|||||.....:|||||.....:|||||.....:
79219 AATGGTCTTACGAGGGAACAGATCTTGCTACTTGTATCGGAAACCG 79268

68 gHisLeuPheGlyGluGluGlyAspArgPheProLeuLeuThrLysI 85
|||||.....:|||||.....:|||||.....:|||||.....:
79269 TGAATTAATTTGGCAATGCTCCAGAACCTGATATTTCCACTTTGACCAAGA 79318

85 IeLeuAspAlaAspGlnAspLeuSerValGlnValHisProAsnAspGlu 101

```

```

|||||..... |||||||||||||||||||
79319 TCCCGAGTCCCAAGACTGCTCAGTGTCCAAATTCACCCAGACGATGCT 79368
|||||..... |||||||||||||||||||
102 TTAAlaasnlleHlsGluasnGlyGluLeuGlyLysThrGluCysTrpTy 118
|||||..... |||||||||||||||||||
79369 TATGAGCTCAGCATGAA...GGCGAACTCGGAAAAAGAAATGCTGTGA 79415
|||||..... |||||||||||||||||||
118 TlellaasPcysGlnLysAspAlaGluIleIleTyrglyHlsAsnAlaT 135
|||||..... |||||||||||||||||||
79416 CATATTCGACCGATGACATGATGATGATGATGATGATGATGATGATGAT 79465
|||||..... |||||||||||||||||||
135 htrHlysgLgluLeuThrThrmellleGluArgGlyLysTrpAspGlu 151
|||||..... |||||||||||||||||||
79466 AGTCAAAAGAAAGAACTCCGCCAAATTCAGGACAAAGAACTGGATGAC 79515
|||||..... |||||||||||||||||||
152 LeuLeuArgArgValLysValLysProGlyAspPhePheTyValProse 168
|||||..... |||||||||||||||||||
79516 TTGTGACCAAAAGTCCCTGTTAAGGCTGAGATTTCTCTATGTACCAAG 79565
|||||..... |||||||||||||||||||
168 TGLYThValHlsAlaIleGlyLysGlyIleLeuAlaLeuGluThrGlu 185
|||||..... |||||||||||||||||||
79566 CGGACTATGACCGCTATCGGGCGGCTATTAATCTTGAATCCAGC 79615
|||||..... |||||||||||||||||||
185 lnaasnsrAspThrThrTyArgLeuTyAspTyArgAspArgLysAspAla 201
|||||..... |||||||||||||||||||
79616 AGTCTACGATACCACTATCGGCTGTATGATGATGATGATGATGATGAT 79665
|||||..... |||||||||||||||||||
202 GluGlyLysLeuArgGluLeuHlsLeuLysSerIleGluValIleG 218
|||||..... |||||||||||||||||||
79666 AAAGCAACTGCGCTGAATCTCATCTTGAATAATCCATGATGCTGTGA 79715
|||||..... |||||||||||||||||||
218 uVal.....ProserIleProGluArgHlsThrValHlsHlsGluGln 233
|||||..... |||||||||||||||||||
79716 CATGTGAGACCCAGCAAAATAGCCGCTCTAATCTGT.....AAAG 79756
|||||..... |||||||||||||||||||
233 legLnaaspleuLeuThrThrThrLeuIleGlyLysAlaTyRheserVal 249
|||||..... |||||||||||||||||||
79757 CAGATGATTTGCGCTCCACCTCTCTGTATCATGATTTCTTCGACGAT 79806
|||||..... |||||||||||||||||||
250 GlySTrPasnLeuSerGlySerAlaSerLeuLysGlnLysProPh 266
|||||..... |||||||||||||||||||
79807 TACAAGTGGGAAATTTACTGGAAGAAAGTGTGAAAGACACAGCTAC 79856
|||||..... |||||||||||||||||||
266 eleuLeuIleSerValIleGluGluGluArgMetIleSerGlyGlu 283
|||||..... |||||||||||||||||||
79857 CAGCTTATGTAGTCTTACTGCTCAAGCTCAAGCTCAAGCTGACAGGA 79906
|||||..... |||||||||||||||||||
283 yValTyRProPheLysGlyAspHlsMetLeuLeuProTyrglyLeu 299
|||||..... |||||||||||||||||||
79907 AAAATTTTCCATTCAAAAGGAGACACTTATCCACCAAGTGTGT 79956
|||||..... |||||||||||||||||||
300 GLyLupPheLysLeuGluGlyTyraL...GluCysIleValSerHls 314
|||||..... |||||||||||||||||||
79957 GAACCTTGACTGCTGGAAGGCAAGGTTGTGAATGATGTGTAGCAT 80004
|||||..... |||||||||||||||||||
seq_name: gb_htg12:AC027136
seq documentation block:
LOCUS AC027136 49897 bp DNA HTG 30-JAN-2001
DEFINITION Staphylococcus aureus clone sabac-106, WORKING DRAFT SEQUENCE, 1
ORDERED PLACES.
ACCESSION AC027136 GI:8567822
VERSION AC027136.5
KEYWORDS HTG; HTGS_PHASE2; HTGS_DRAFT.
SOURCE Staphylococcus aureus.
ORGANISM Staphylococcus aureus
Bacteria; Firmicutes; Bacillus/Clostridium group;
Staphylococcus group; Staphylococcus.
1 (bases 1 to 49897)
REFERENCE 1. Iohn, P., Ol.S., Ray, L., Ford, B., Iondola, J., and Roe, B. A.
AUTHORS Iohn, P., Ol.S., Ray, L., Ford, B., Iondola, J., and Roe, B. A.
TITLE Direct Submission
JOURNAL Submitted (28-MAR-2000) Department of Chemistry And Biochemistry,
The University of Oklahoma, 620 Parrington Oval, Room 208, Norman,
OK 73019, USA
COMMENT On Jan 15, 2000 this sequence version replaced gi:8225121.
* NOTE: This is a 'working draft' sequence. It currently
* consists of 1 contigs. Gaps between the contigs
* are represented as runs of N. The order of the pieces
* is believed to be correct as given, however the sizes
* of the gaps between them are based on estimates that have
* been provided by the submitter.
* This sequence will be replaced
* by the finished sequence as soon as it is available and
* the accession number will be preserved.
1 49897: contig of 49897 bp in length.
* Location/Qualifiers
FEATURES
1. 49897
source
BASE COUNT 15753 a 9237 c 7267 g 17640 t
ORIGIN
alignment_scores:
Quality: 819.50 Length: 312
Ratio: 3.415 Gaps: 1
Percent Similarity: 76.923 Percent Identity: 50.321
alignement block:
US-09-462-846-2 x AC027136 ..
Align seg 1/1 to: AC027136 from: 1 to: 49897
3 ThrGluProLeuPhePheLysProValPheLysGluArgIleTrpGly 19
|||||..... |||||||||||||||||||
35576 ACTATGGCATATTTTACAAACCCGTTTAAAGACGAATTTGGGCGG 35625
|||||..... |||||||||||||||||||
19 yThAlaLeuAlaasPheGlyTyThrIleProSerGlnArgTyrgly 36
|||||..... |||||||||||||||||||
35626 TCATGACATTAATAAGCATTTTACATGATATCCCAATGAACAACAGTG 35675
|||||..... |||||||||||||||||||
36 lucSTrPAlaPheAlaHlsGlnasnGlyGlnSerValIleGlnAsn 52
|||||..... |||||||||||||||||||
35676 AATGCTGGCAATTTCTGCACATCCAAATAGTCTTATTAATTAATCAAT 35725
|||||..... |||||||||||||||||||
53 GlyMetTyrglyGlyPheThrLeuSerGluLeuTrpGluHlsArgH 69
|||||..... |||||||||||||||||||
35726 GGTCCATACCAAGATATGACCCATGATCAACTTGTGTCACACATGCTGA 35775
|||||..... |||||||||||||||||||
69 sleuPheGlyLnuLeuGluLysAspArgPheProLeuLeuThrLysIle 86
|||||..... |||||||||||||||||||
35776 ATTAATCGATATGATTCACAGATAGTTCCTCCCTGCTAATCAAGTAT 35825
|||||..... |||||||||||||||||||
86 euasPAlaasPLeuPheLysSerValGlnValHlsProAsnAspGluTy 102
|||||..... |||||||||||||||||||
35826 TAGATGCCAATGATTAATTAATCTTCAAGTCAACCCAGATGATGAC 35875
|||||..... |||||||||||||||||||
103 AlaasnlleHlsGluasnGlyGluLeuGlyLysThrGluCysTrpTy 119
|||||..... |||||||||||||||||||
35876 GCTTAAACAGAA...GGCGAACTGAGTAAACACAAATGTTGGTATAT 35922
|||||..... |||||||||||||||||||
119 eileasPcysGlnLysAspAlaGluIleIleTyrglyHlsAsnAlaTh 136
|||||..... |||||||||||||||||||
35923 TTATGATGACCAAGGCTGCAATTAATATATATGCGGTATGATATAT 35972
|||||..... |||||||||||||||||||
136 htrVsgLgluLeuThrThrmellleGluArgGlyLysTrpAspGluLe 152
|||||..... |||||||||||||||||||
35973 ATAAACAGACTTATAGACATGATTAATCAATGATTTGACAGCTT 36022
|||||..... |||||||||||||||||||
153 LeuArgArgValLysValLysProGlyAspPhePheTyValProser 169

```

```

AUTHORS Iohn, P., Ol.S., Ray, L., Ford, B., Iondola, J., and Roe, B. A.
TITLE Direct Submission
JOURNAL Submitted (28-MAR-2000) Department of Chemistry And Biochemistry,
The University of Oklahoma, 620 Parrington Oval, Room 208, Norman,
OK 73019, USA
COMMENT On Jan 15, 2000 this sequence version replaced gi:8225121.
* NOTE: This is a 'working draft' sequence. It currently
* consists of 1 contigs. Gaps between the contigs
* are represented as runs of N. The order of the pieces
* is believed to be correct as given, however the sizes
* of the gaps between them are based on estimates that have
* been provided by the submitter.
* This sequence will be replaced
* by the finished sequence as soon as it is available and
* the accession number will be preserved.
1 49897: contig of 49897 bp in length.
* Location/Qualifiers
FEATURES
1. 49897
source
BASE COUNT 15753 a 9237 c 7267 g 17640 t
ORIGIN
alignment_scores:
Quality: 819.50 Length: 312
Ratio: 3.415 Gaps: 1
Percent Similarity: 76.923 Percent Identity: 50.321
alignement block:
US-09-462-846-2 x AC027136 ..
Align seg 1/1 to: AC027136 from: 1 to: 49897
3 ThrGluProLeuPhePheLysProValPheLysGluArgIleTrpGly 19
|||||..... |||||||||||||||||||
35576 ACTATGGCATATTTTACAAACCCGTTTAAAGACGAATTTGGGCGG 35625
|||||..... |||||||||||||||||||
19 yThAlaLeuAlaasPheGlyTyThrIleProSerGlnArgTyrgly 36
|||||..... |||||||||||||||||||
35626 TCATGACATTAATAAGCATTTTACATGATATCCCAATGAACAACAGTG 35675
|||||..... |||||||||||||||||||
36 lucSTrPAlaPheAlaHlsGlnasnGlyGlnSerValIleGlnAsn 52
|||||..... |||||||||||||||||||
35676 AATGCTGGCAATTTCTGCACATCCAAATAGTCTTATTAATTAATCAAT 35725
|||||..... |||||||||||||||||||
53 GlyMetTyrglyGlyPheThrLeuSerGluLeuTrpGluHlsArgH 69
|||||..... |||||||||||||||||||
35726 GGTCCATACCAAGATATGACCCATGATCAACTTGTGTCACACATGCTGA 35775
|||||..... |||||||||||||||||||
69 sleuPheGlyLnuLeuGluLysAspArgPheProLeuLeuThrLysIle 86
|||||..... |||||||||||||||||||
35776 ATTAATCGATATGATTCACAGATAGTTCCTCCCTGCTAATCAAGTAT 35825
|||||..... |||||||||||||||||||
86 euasPAlaasPLeuPheLysSerValGlnValHlsProAsnAspGluTy 102
|||||..... |||||||||||||||||||
35826 TAGATGCCAATGATTAATTAATCTTCAAGTCAACCCAGATGATGAC 35875
|||||..... |||||||||||||||||||
103 AlaasnlleHlsGluasnGlyGluLeuGlyLysThrGluCysTrpTy 119
|||||..... |||||||||||||||||||
35876 GCTTAAACAGAA...GGCGAACTGAGTAAACACAAATGTTGGTATAT 35922
|||||..... |||||||||||||||||||
119 eileasPcysGlnLysAspAlaGluIleIleTyrglyHlsAsnAlaTh 136
|||||..... |||||||||||||||||||
35923 TTATGATGACCAAGGCTGCAATTAATATATATGCGGTATGATATAT 35972
|||||..... |||||||||||||||||||
136 htrVsgLgluLeuThrThrmellleGluArgGlyLysTrpAspGluLe 152
|||||..... |||||||||||||||||||
35973 ATAAACAGACTTATAGACATGATTAATCAATGATTTGACAGCTT 36022
|||||..... |||||||||||||||||||
153 LeuArgArgValLysValLysProGlyAspPhePheTyValProser 169

```


36023 TACAAACGATACCGCTTAAAGCCGGGTACCTTTATATGTTCCGCTGG 36072
 169 yThrValHisAlaIleGlyIysGlyIleLeuAlaLeuLuhThrGlnGlnA 186
 36073 CACTGTCCATGCCATGTGGTCCGGATATTTTAATTTTAAGACACAAACAT 36122
 186 snSerAspThrThrThyValArgLeuThyrAspTyrAspArgLysAspIaGlu 202
 36123 CCTCCGATACGACAAACCGTATTTATGACTAATGACAGACGATCAAAAT 36172
 203 GlyIysLeuArgGlnLeuHisLeuLysSerIleGluValIleGluVal 219
 36173 GGTCATTGCCGATTTGGCATTGACATTACAGCAAAATAAAGATGATTTGATT 36222
 219 IProSerIleProGluValGlnHisThrValHisIleGlnIleGluAspL 236
 36223 AGGCATCATGTGAACCAAAATACACTCCGTATAAACAATCACAACGACAC 36272
 236 euLeuThrThrThrLeuIleGluCysAlaTyrPheSerValGlyLysTrp 252
 36273 ACACGTCCATCAATTTTGTAGATATATCATTAATTTTGCAGTAAATATAAATGG 36322
 253 AsnIleSerGlySerAlaSerLeuLysGlnGlnLysProPheLeuLeuI 269
 36323 GATATTACAGAGATTTAAATATTCATTAATCAACACATGCTTATGCTTACT 36372
 269 eSerValIleGluGlyGlyIArgMetIleSerGlyLysIuTyValTyrP 286
 36373 TACAAATTAACGATGTACGACGATCTTAAACAGTAACGATAAGCACTAAG 36422
 286 rOPheIysGlyAspHisMetLeuLeuProTyrGlyLeuAlaGluPhe 302
 36423 ACATCCAAAAGGCTTACTTATTTTAACTACGGAAGACAGAGATATT 36472
 303 LysLeuGlyGlyTyrAlaGluLysIleValSerHis 314
 36473 CACTTGTGGGTAAATTTATGTGTATAGTAAGTCAI 36508


```

166 valPProserGlyThrValHisAla11IleGlyValGlyL1IleuAla1LeuCl 182
167 ::::::::::::::::::::::::::::::::::::::::::::::::::::::::::::
193 ATTCCGACGACGACATCCACTGCTGCGGCGCTCAACACCTGCTGTGTGA 144
182 uThrGlnGlnHisSerAspThr...ThrTrpArgLeuTyraPtyrAspA 198
143 AATCAGCGCTACGCCCTACATATTTCACCTTCACAGATGTTGATGCTGATGC 94
198 TGLYSALAGLagLgLYLSLeuArgL1IleuHisLeuLYSLysSerIle 214
111 ::::::::::::::::::::::::::::::::::::::::::::::::::::::::::::
93 GCCTTGACCTGTGATGAGACGCCCGGCGCTCAACATCGACGCGCTTC 44

215 Gtu 215
111
43 GAG 41

seq_name: gb-gss2:AQ158357

seq_documentation_block:
LOCUS      AQ158357                655 bp            DNA             12-SEP-1998
DEFINITION nxbdb011003r CUG1 Rice BAC Library Oryza sativa genomic clone
ACCESSION  nxbdb011003r, DNA sequence.
VERSION    AQ158357
KEYWORDS   AQ158357.1 GI:3555382
SOURCE     GSS.
ORGANISM   Oryza sativa.
            Oryza sativa
            Eukaryota; Viridiplantae; Embryophyta; Tracheophyta; Spermatophyta
            Magnoliophyta; Liliopsida; Poales; Poaceae; Ehrhartoideae; Oryzaeae
            Oryza.
            1 (bases 1 to 655)
            Wang, R.A. and Dean, R.A.
            A BAC End Sequencing Framework to Sequence the Rice Genome
            Unpublished (1998)
            Contact: Ming RA
            Clemson University Genomics Institute
            Clemson University
            100 Jordan Hall, Clemson, SC 29654, USA
            Tel.: 864 656 7288
            Fax: 864 656 4293
            Email: rtwing@clemson.edu
            Seq primer: GGAAACAGCTATGACATG
            Class: BAC ends
            High quality sequence stop: 398.
            Location/Qualifiers
                1..655

```

BASE COUNT	124 a	212 c	212 g	107 t	ORIGIN
alignment_scores:					
Quality:	118.50			Length: 164	
Ratio:	1.539			Gaps: 11	
Percent Similarity:	46.951			Percent Identity: 32.317	
alignment_block:					
us-09-462-846-2 x A0158357	..				
Align seg 1/1 to: A0158357 from: 1 to: 655					
9	lysprvalaiphelysglaargylletrpglyuglyutpralaleuilaaprph	25			
201	AAACAATCTTTTCGGGGCGGATGCGGGAGATCGGGAGCCGCGATCGCCACCA	250			
25	e.....				
251	TCGCGTCGCGCTTCGCGTCGCGTCGCGTCGCGTCGCGTCGCGTCGCGTCGCG	300			
32	glatgtrhlglylucystpralaparleaalaaniagslanaglylase	48			
301	ACG.....				
48	valvalaglanaglymetyludysglpether.....	59			
333	CTGGGCGCGCCACACGCTGTATCGGGCTTCGCGCGATCGCCGCAAGATG	382			
60Leusergluleutprglinhisatgihisaleuphegllyglh	73			
383	GGGATCGGTCGGCGAGGCTCG.....	411			
74	leuglulglyasprapgrphero.....leuautrhytyleuaspalaas	89			
412	ACGCGCGCGATGTCACCGCGGACGCTGTGATCGATGTGTTGCACAG	461			
89	pclnasprleuservalaglnvalaisproasnsprlyutrylalaanlleh	106			
462	CGAGAAATGTTCGGTCGAGGTCACACCGAAGATGAGCAGCG.....C	505			
106	lsglanagly.....glnleuglylgtthrghlucysatprtyrllleaspr	121			
506	ACGCGCGCGGCTCGCGCGCGCAAGAGATGCTGGTATGCTCCATAC	555			
122	cysgln.lyaspralagluillelelygly.....hisnalathrhl	137			
556	CCCCAAGCCGATTCACGATCGCGCTGCGCCAAACAAACCCATATGCC	605			
137	ysglugluleutrthhmettleaglucgilyglutpr	149			
606	GGGAGGAACTTCG.....GGCGGGCGGCTTCG	654			
seq_name: gb_est29:AL537077					
seq_documentation_block:					
LOCUS	AL537077	830 bp	mRNA	EST	13-FEB-2001
DEFINITION	AL537077 ltr1.fl013.fbrn1 Homo sapiens cDNA clone CS0DF019YF01.5				
ACCESSION	AL537077				
VERSION	AL537077.1	GI:12800570			
KEYWORDS	EST.				
SOURCE	human.				
ORGANISM	Homo sapiens				
REFERENCE	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;				
AUTHORS	Mammalia; Eutelesta; Primates; Catarrhini; Homnidae; Homo.				
TITLE	1 (bases 1 to 830)				
JOURNAL	Li, W.B., Gruber, C., Jesse, J. and Polayes D.				
COMMENT	Full-length cDNA libraries and normalization				
	unpublished (2001)				
	Contact: Genoscope				
	Genoscope - Centre National de Sequencage				

ORIGIN

alignment_scores: Quality: 93.00 Length: 200
 Ratio: 0.959 Gaps: 8
 Percent Similarity: 48.500 Percent Identity: 24.000

alignment_block:
 US-09-462-846-2 x BF492790 ..

Align seg 1/1 to: BF492790 from: 1 to: 675

```

146 ArgGlyGluTrpAspGluLeuLeuArgValysValysProGlyAs 162
|||||                   ::: |||||                   |||:::
51 CAGTGCATCAACGACGACACCTGAGAGCATTT.....GGCAA 88
162 phePheTyrValProSerGlyThrValHisAlaIle.....GlyLysG 177
:::||||                   :::||||                   |||
89 CAGTGTTCGACTCTCTCCGCCGCCGAGCAAGTTGAGACACACAAAG 138
177 LylLeuAlaLeuGluThrGlnGlnAsnSerAspThrThrTyrArgLeu 193
|||||                   ::: |||||                   :::
139 GCATCCTTCGATACATGCGCCACGACGACGACAGATGTGCAGACAAAGTG 188
194 TyrAspTyrAspArgLysAspAla.....Gly 202
|||||                   :::||||                   ||
189 GCGGACTGCTGCTGGCGGAGGCTTCGCTCGGTTGCTCTTAAAGGA 238
202 uGlyLysLeuArgGluLeuHisLeuLysLysSerIleGluValIleGluV 219
|||||                   :::||||                   :::
239 GGGG.....ATCCACGTTGCACTATCCGACGAGTGATGGAGC 276
219 aLProSerIleProGluArgHisThrValHisHisGluGlnIleGluAsp 235
:::                   :::|||||                   ||| |||
277 GCGGCACCTTTTGGCACCGGACACGACACGCTCTGACACACAGTCGAGAAC 326
236 LeuLeuThrThrThrLeuIleGluCysAlaTyr.....Ph 247
:::                   :::                   |||
337 AAGGTGGGTGACAACTCACTGACACACCTATCCGACACAGGCGCCCTA 376
247 eSerValGlyLysTrpAsnLeuSerGlySerAlaSerLeuLysGlnGlnI 264
|||||                   ::: |||||                   |||
377 CTCGCTCTCCAAACAGTCCCTCTGCGAGTGTGGCGCTCGGCTTCGAGC 426
264 ySPhePheLeuLeuIleSer.....ValIleGluGlyGluGlyArg 277
:::                   |||                   :::
427 ACGGCTACTCCATGGCCAGTCCCATGCTCTGTCATGTGTGGAGAGCCAG 476
278 .....MetI 279
477 TTTGGGACCTTTCGCAACAGCGGCGCATGATCATGACAGCTTCATAGC 526
279 eSerGlyGluTyrValTyrProPheLysGlyAspHisMetLeuLeuP 296
|||||                   :::                   |||||
527 CAGCGGCGAACAAGTGGGTGGCGGACGTCGAGATCTGCTGTCGCGC 576
296 rGlyr.....GlyLeuGlyGluPheLysLeuGluGlyTyrAlaGlu 309
|||||                   |||                   :::
577 CACACACACATGAGAGGCGATGGGCGCGGACGACCTTTCGCGCGCATGAG 626
seq_name: gb_estf5:BE567646
seq_documentation_block:   869 bp                   EST                   15-AUG-2000
LOCUS   BE567646                   mRNA                   IMAGE:3682654 5',
DEFINITION   6013402933F1 NIH_MGC_53 Homo sapiens cDNA clone IMAGE:3682654 5',
ACCESSION   BE567646                   EST
VERSION   BE567646.1                   GI:9811366
KEYWORDS   EST.
SOURCE   human.
ORGANISM   Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

```

REFERENCE
 1 (bases 1 to 869)
 AUTHORS NIH-MGC <http://mgc.nci.nih.gov/>.
 TITLE National Institutes of Health, Mammalian Gene Collection (MGC)
 JOURNAL Unpublished (1999)
 CONTACT Robert Strausberg, Ph.D.
 Tel: (301) 496-1550
 Email: Robert.Strausberg@nih.gov

FEATURES
 source
 1. 869
 /organism="Homo sapiens"
 /db_xref="taxon:9606"
 /clone="IMAGE:3682654"
 /clone_id="NIH_MGC_53"
 /tissue_type="carcinoma, cell line"
 /lab_host="DH10B (T1 phage-resistant)"
 /note="Organ: bladder; Vector: pDMR-LIB (Clontech); Site_1: SfiI (ggcgcctcggcc); Site_2: SfiI (ggcgcataagcc 5' and 3' adaptors were used in cloning as follows: 5' adaptor sequence: 5'-CACGCCATATGACC-3' and 3' adaptor sequence: 5'-ATCTGAGAGCCGACGCGCCGACATG-dt(30)BN-3' (where B = A, C, or G and N = A, C, G, or T). Average insert size 1.55 kb (range 0.9-4.0 kb). 15/15 colonies contained inserts by PCR. This library was enriched for full-length clones and was constructed by Clontech Laboratories (Palo Alto, CA)."
 High quality sequence stop: 481.
 Location/Qualifiers

BASE COUNT 196 a 228 c 215 g 230 t

alignment_scores: Quality: 93.00 Length: 241
 Ratio: 0.809 Gaps: 11
 Percent Similarity: 47.718 Percent Identity: 20.332

alignment_block:
 US-09-462-846-2 x BE567646/rev ..

Align seg 1/1 to reverse of: BE567646 from: 1 to: 869

```

58 PheThrLeuSerGluLeuTrpGluHisArgHisLeuPheGlyGln 74
:::                   :::|||||                   :::
833 TATGATATAGCTGTGATGTGGCACCCGCC.....TCATGAGA 796
74 uGluGlyAspArgPheProLeuLeuThrLysIleLeuAspAlaSerI 91
|||||                   :::                   |||
795 AGAGAGGGGCTTTTTCACACAGACACAGATGTGTGGCGCGCCGCC 746
91 sPheSerValGlnValHisProAsnAspGluTyrAlaAsnIleGlu 107
:::                   |||                   |||||
745 GGGGCCGCCCTTACGTGTCGCA.....CAGGAA 717
108 AsnGlyGluLeuGlyThrGluCysTrpTyrIleIleAspCysGln 124
:::                   |||                   |||||
716 CAACACACGTGGCGGCTTTGATGTT.....TCGATATA 682
124 sAspAlaGluIleIleTyrGlyHisAsnAlaThrThrLysGluGluLeu 141
|                   :::                   |||||
681 A...AAGAAATGCCATTAAGCGTGGGCGCCACCAATATTAACGAGTTA 635
141 hTrhMetIleGlu.....ArgGlyGluTrpAspGluLeuLeuArg 155
|||||                   ::: |||||                   :::
634 CCGTTTATTGACCCACAAAGAAATGTGGGAAG..... 597

```


FBan0079340 last_updated:000321, mRNA sequence.
 ACCESSION BF489340
 VERSION BF489340.1 GI:11572641
 KEYWORDS EST
 SOURCE fruit fly
 ORGANISM Drosophila melanogaster
 Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
 Pterygota; Neoptera; Endopterygota; Diptera; Brachycera;
 Muscomorpha; Ephydroidea; Drosophilidae; Drosophila.
 REFERENCE 1 (bases 1 to 577)
 Stapleton, M., Brokstein, P., Hong, L., Agbayani, A., Baxter, E., Berman
 B., Carlson, J., Champs, M., Chavez, C., Chew, M., Dorsett, V., Farfan
 D., Frise, E., George, R., Gonzalez, M., Guarino, H., Harris, N., Li, P.,
 Liao, G., Miranda, A., Mirra, S., Mungall, C. J., Nunoo, J., Pacleb, J.,
 Park, S., Paragas, V., Phouanavong, S., Wan, K., Xu, C., Celniker, S.,
 Lewis, S. E. and Rubin, G. M.
 TITLE Berkeley Drosophila Gene Collection Project
 JOURNAL Unpublished (2000)
 COMMENT Contact: Stapleton, M.
 BDGP
 Lawrence Berkeley National Lab
 One Cyclotron Rd, Berkeley, CA 94720, USA
 Fax: 510 486 6798
 Email: <http://www.fruitfly.org/EST>, est@fruitfly.berkeley.edu
 hit genomic sequence AE003525; hit genomic sequence AE003472
 Plate: AT 253 row: C column: 2
 High quality sequence stop: 575.
 Location/Qualifiers
 1..577
 /organism="Drosophila melanogaster"
 /db_xref="taxon:7227"
 /clone="AT25326"
 /clone_1lb="AT Drosophila melanogaster adult testes potb7"
 /sex="male"
 /dev_stage="0-3 day old Ore-R males"
 /lab_host="DMS-alpha or DMS-alpha Tona as per database (AT
 121 on are in Tona cells)"
 /note="Organ: ADULT testes; Vector: potb7; Site: 1: EcoRI;
 Site 2: XhoI; The mRNA for the testis library was made
 from testes and seminal vesicles hand dissected from 0-3
 day old Ore-R males. RNA kindly provided by the lab of
 Margaret Fuller. Sized fractionated cDNAs were directly
 ligated into potb7. Plasmid cDNA library."
 BASE COUNT 114 a 187 c 171 g 105 t
 ORIGIN
 alignment_scores: Length: 185
 Quality: 92.00 Gaps: 7
 Ratio: 1.011 Percent Identity: 24.324
 Percent Similarity: 49.189
 Alignment block:
 US-09-462-846-2 x BF489340 ..
 Align seg 1/1 to: BF489340 from: 1 to: 577
 161 GlyasphepheyValProserglythryValhisalale.....GI 175
 |||:|||||: |||:
 32 GGCACACATGTTCTCGACCTCCGCCGCCGACACAGTTCGAGACACA 81
 |||:|||||: |||:
 175 YLYSGLYLELeuAlaLeuGIuThrGlnGlnAsnseraspthrThrTyra 192
 |||:|||||: |||:
 82 CAAGGGCATCTTCGCATACGTGCCACGACGACAGATGTGACAGACA 131
 |||:|||||: |||:
 192 rglauTyraPtyrAsparGlyAspAla.....201
 |||:|||||: |||:
 132 AAGGGGGGACGTGCTGGCGGCGGCGCTTCGCCCTTCGCTCCTT 181
 |||:|||||: |||:
 202 ...GluGlyLysLeuArgGluLeuHisLeuLysSerIleGluVal11 217
 |||:|||||: |||:
 182 AAGGAGGCGC.....ATCCACGCTTGACATACCGGCCGACGATGT 219

217 eglValProserIleProGluArgHisThrValHisleGIuGln11eg 234
 |||:|||||: |||:
 220 GGAGCGCGGACACCTTTTGCACCGGACGACGACCTTCGACACGCG 269
 |||:|||||: |||:
 234 lnaSpLeuLeuThrThrThrIleuIleGIuGlySalATyr.....246
 |||:|||||: |||:
 270 AGGACAGAGGGGTGTCACACTGACGACACCTCTTATCCGACACGCG 319
 |||:|||||: |||:
 247 ...PhseValGIuLysTrpAsnLeuSerGIySerAlaSerLeuysgl 262
 |||:|||||: |||:
 320 CCTACTCCGCTCCACACAGCTCCTGTGCGAGTGTCCGCTTCGCTT 369
 |||:|||||: |||:
 262 ngInLysProPheLeuLeuIleSer.....ValIleGIuGIuGln 276
 |||:|||||: |||:
 370 CGAGCAGGCGCTACTCCATGCGCAGTCCCAATGCTGTGTATGTGGAG 419
 |||:|||||: |||:
 276 lYArg.....277
 |||:|||||: |||:
 420 GACAGTTGGGGACTTCGACACAGCGCGCAGTCATCGACAGCTTC 469
 |||:|||||: |||:
 278 MetIISerGIuGIuTyraValTyrrProPheLysGIuAspHisMetIe 294
 |||:|||||: |||:
 470 ATAGCCAGCGCGGAAACCAAGTGGGTGGCGGACGTCGAGTCGTATG 519
 |||:|||||: |||:
 294 ulauProTyr.....GlyLeuGIuGluPheLysLeuGIuGIuTyra 308
 |||:|||||: |||:
 520 GGTGCCACACACATGAGAGGCGATGGCCCTTGACGACTTCCGCGCGA 569
 |||:|||||: |||:
 308 laGIu 309
 |||:
 570 TTGAG 574
 seg_name: gb_est71:BE204501
 seg_documentation block:
 LOCUS BE204501 mRNA EST 05-SEP-2000
 DEFINITION EST397177 KVO Medicago truncatula cDNA clone PKV0-16010, mRNA
 sequence.
 ACCESSION BE204501
 VERSION BE204501.1 GI:8747787
 KEYWORDS EST.
 SOURCE barrel medic.
 ORGANISM Medicago truncatula
 Eukaryota; Viridiplantae; Embryophyta; Tracheophyta; Spermatophyta;
 Magnoliophyta; eudicotyledons; core eudicots; rosidae; eurosids I;
 Fabales; Fabaceae; Papilionoideae; Medicago.
 1 (bases 1 to 551)
 Vandenbosch, K., Endre, G., Hur, J., Moore, J., Beremand, P., Ellis, L.,
 Town, C.D., Bowman, C.L., Craven, M.B., Hansen, T.S., Holt, I.E. and
 Fraser, C.M.
 ESTs from uninoculated seedling roots of Medicago truncatula
 Unpublished (1999)
 TITLE JOURNAL
 COMMENT Contact: Vandenbosch K
 Department of Biology
 Texas A&M University
 College Station, TX 77843-3258, USA
 Tel: 409 845 7707
 Fax: 409 845 2891
 Email: kate@mail.bio.tamu.edu
 Texas A&M University name: T264273e
 TIGR sequence name: MNCAGV89TK
 More information is available at:
<http://chrystle.tamu.edu/medicago>
 Seq primer: Skmod (CTA gAA CTA gTg gAT CC).
 Location/Qualifiers
 1..551
 /organism="Medicago truncatula"
 /cultivar="genotype A17"
 /db_xref="taxon:3880"
 /clone="PKV0-16010"
 /clone_1lb="KV0"
 /tissue_type="Seedling roots"
 /dev_stage="Immediately prior to inoculation with

Page 8

seq_name: gb_est88:BF497790

VERSION BF497790.1 GI:11581091

fruit fly:
Drosophila melanogaster

AUTHORS
 Stalleon, M., Brokstein, P., Hong, L., Agbayan, A., Baxter, E., Berman, B., Carlson, J., Champe, M., Chavez, C., Chew, M., Dorsett, V., Farfan, D., Frise, E., George, R., Gonzalez, M., Gunn, H., Harris, N., Li, P., Ilaio, G., Miranda, A., Misra, S., Mungell, C. J., Nunoo, J., Pacled, J., Park, S., Paragas, V., Phoumenavong, S., Wan, K., Yu, C., Zellmeyer, S.

TITLE	Berkeley Drosophila Gene Collection Project
JOURNAL	Unpublished (2000)
COMMENT	Contact: Stapleton, M. 1998

Lawrence Berkeley National Lab
One Cyclotron Rd. Berkeley, CA 94720, USA
Fax: 510 486 6798
Email: http://www.fruitfly.org/EST_esa@fruitfly.berkeley.edu
hit_genomic.sequence AB005525; hit_genomic.sequence AB0053472
Plate: A1:120 row: G column: 10
High quality sequence stop: 598.

Location/Qualifiers
1. .668

```

/organism="Drosophila melanogaster"
/db_xref="taxon:7227"
/clone="Ari1082"
/clone_1fb="Ari Drosophila melanogaster adult testes potB7
/sex="male"
/dev_stage="0-3 day old Ore-R males"
/lab_host="DHS-alpha or DHS-alpha Tona as per database (Ari
121 on are in Tona cells)"
/note="Organ: Adult testes; Vector: potB7; Site_1: EcoRI;
Site_2: XhoI; The mRNA for the testis library was made
from testes and seminal vesicles hand dissected from 0-3
day old Ore-R males. RNA kindly provided by the lab of
Margaret Fuller. Sized fractionated cDNAs were directly
ligated into potB7. Plasmid cDNA library."

```

alignment_scores:		
Quality:	91.50	Length: 193
Ratio:	0.984	Gaps: 8
Percent Similarity:	48.187	Percent Identity: 23.834

Align seg 1/1 to: BF497790 from: 1 to: 668

[illegible]

seq_name: gb_est88:BF491677

seq. documentation, block:			
LOCUS	BFA91677	708 bp	mRNA EST 06-DEC-2000
DEFINITION	A178475 5prime AT Drosophila melanogaster adult testes polyB7		
	Drosophila melanogaster cDNA clone A178475 5 similar to C61161		
	Phan001661 lasL updated:000321, mRNA sequence.		

ACCESSION	BF491677	GI:11574978
VERSION	BF491677.1	
KEYWORDS	EST.	
SOURCE	Fruit fly.	

ORGANISM	<i>Drosophila melanogaster</i> Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta; Pterygota; Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha; Ephydroidea; Drosophilidae; <i>Drosophila</i> . 1 (bases 1 to 708)
REFERENCE	Stapleton, M., Broksstein, P., Hong, L., Agbayani, A., Baxter, E., Berman
AUTHORS	

Stapleton, C.M., Brocksiehn, P., Hong, L., Agbayan, A., Baxter, E., Berman, B., Carlson, J., Champe, M., Chavez, C., Chew, M., Dorsett, V., Farfan, D., Frisze, E., George, R., Gonzalez, M., Guarin, H., Harris, N., Li, P., Liao, G., Miranda, A., Mitra, S., Mungall, C. J., Nunoo, J., Pacleb, J.,

Park, S., Paragas, V., Phouanavong, S., Wan, K., Yu, C., Celniker, S.,
Lewis, S.E. and Rubin, G.M.
Berkeley Drosophila Gene Collection Project
Unpublished (2000)
COMMENT
Contact: Stapleton, M.
BDGP

Lawrence Berkeley National Lab
One Cyclotron Rd, Berkeley, CA 94720, USA
Fax: 510 486 6798
Email: http://www.fruitfly.org/EST_estefruitfly.berkeley.edu
hit genomic sequence AE003525; hit genomic sequence AE003472
Plate: AT.264 row: 6 column: 3
High quality sequence stop: 681.
Location/Qualifiers

FEATURES

source

1. 708
/organism="Drosophila melanogaster"
/db_xref="taxon:7227"
/clone_lib="AT28475"
/sex="male"
/dev_stage="0-3 day old Ore-R males"
/lab_host="DH5-alpha or DH5-alpha Tona as per database (AT
121 on are in Tona cells)"
/note="Organ: Adult testes; Vector: pOTB7; Site_1: EcoRI;
Site_2: XhoI; The mRNA for the testis library was made
from testes and seminal vesicles hand dissected from 0-3
day old Ore-R males. RNA kindly provided by the lab of
Margaret Fuller. Sized fractionated cDNAs were directly
ligated into pOTB7. Plasmid cDNA library."
BASE COUNT 148 a 227 c 203 g 130 t
ORIGIN

alignment_scores:
Quality: 91.50 Length: 193
Ratio: 0.984 Gaps: 8
Percent Similarity: 48.187 Percent Identity: 23.834

alignment_block:
US-09-462-846-2 x BF491677 ..

Align, seg 1/1 to: BF491677 from: 1 to: 708

```

155 ArgValLysValLysPro..... 160
156 |||:|||||:|||||
157 CGCTGAACCTGTGCCCTGCATGACGACGACACCTGACAGCAT 174
158 |||:|||||:|||||
159 GAspPhePhePheTyrValProSerGlyThrValHisAlaIle.....G 175
160 |||:|||||:|||||
161 TGCGACACATGTTCTCGACTCTCGCGCGCGGACGACAGTTCGAGAC 224
162 |||:|||||:|||||
163 LysGlyIleLeuAlaLeuGlnThrGlnGlnInsSerAspThrTyr 191
164 |||:|||||:|||||
165 ACAAGGCACTCTTCGCACTGCGCCAGCGACGACGATGTGCGAGAC 274
166 |||:|||||:|||||
167 ArgLeuTyrAspTyrAspArgLysAspAla..... 201
168 |||:|||||:|||||
169 AAAGTGGCGACTGTGCTGCGGCGAGGCTTCGCTCGCTCGCTCT 324
170 |||:|||||:|||||
171 TAAGGAGAGGC.....ATCCAGCTTCGACATATCCGCGCAGGATG 362
172 |||:|||||:|||||
173 IeGluAlaProSerIleProGluArgHisThrValHisHisGluGlnIle 233
174 |||:|||||:|||||
175 TGAAGCGCGGACCTTTCGACCGGACGACGATCTCTGACACGACGTCG 412
176 |||:|||||:|||||
177 GluAspLeuLeuThrThrThrLeuIleGluCysAlaTyr..... 246
178 |||:|||||:|||||
179 GAGGACAGAGTGTGTACACTGACCTGACCTGATCCGACGACGAC 462
180 |||:|||||:|||||
181 PheSerValGlyLysTyrPasnLeuSerGlySerAlaSerLeuLysG 262
182 |||:|||||:|||||

```

```

463 GCCTCTCTCCGCTCTCCACAGCTCCCTGTGCGAGTGTGGGCTTGTGCT 512
464 |||:|||||:|||||
465 InGlnLysProPheLeuLeuIleSer.....ValIleGlnGlyGlu 275
466 |||:|||||:|||||
467 TCAGGACGCGCTACTTCATGTCAGTCCCATGCTCTGTGTCATGTGGAA 562
468 |||:|||||:|||||
469 GlyArg..... 277
470 |||:|||||:|||||
471 GGACACTTGGGAGACTTCTGCAACAGCGGCGCATGTCATGACACGTT 612
472 |||:|||||:|||||
473 MetIleSerGlyGlyTyrValTyrProPheLysLysGlyAspHisMetL 294
474 |||:|||||:|||||
475 CATAGCGACGCGGCAACCAAGTGTGCGCGCATGTCGATGT 652
476 |||:|||||:|||||
477 euleuProTyr.....GlyLeuGly 300
478 |||:|||||:|||||
479 TGCTGCCACACAGCATGTGAGGCGCATGTGGC 691
480 |||:|||||:|||||
481 seq_name: gp_est88:BF488952

```

seq_documentation_block: 652 bp mRNA EST 06-DEC-2000
LOCUS BF488952
DEFINITION AT24629.5prime AT Drosophila melanogaster adult testes pOTB7
Drosophila melanogaster cDNA clone AT24629 5 similar to CG11661:
Fban0011661 last updated:000321, mRNA sequence.
ACCESSION BF488952
VERSION BF488952.1 GI:11572253
KEYWORDS EST.
SOURCE fruit fly.
ORGANISM Drosophila melanogaster
Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
Pterygota; Neoptera; Endopterygota; Diptera; Brachycera;
Muscomorpha; Ephydroidea; Drosophilidae; Drosophila.

REFERENCE
1 (bases 1 to 652)
Stapleton, M., Brokstein, P., Hong, L., Asbayani, A., Baxter, E., Berman
B., Carlson, J., Champe, M., Chavez, C., Chew, M., Dorsett, V., Farfan
D., Frise, E., George, R., Gonzalez, M., Guarin, H., Harris, N., Li, P.,
Liao, G., Miranda, A., Misra, S., Mungall, C.D., Nunoo, J., Pacleb, D.,
Park, S., Paragas, V., Phouanavong, S., Wan, K., Yu, C., Celniker, S.,
Lewis, S.E. and Rubin, G.M.
Berkeley Drosophila Gene Collection Project
Unpublished (2000)
JOURNAL
COMMENT
BDGP
Lawrence Berkeley National Lab
One Cyclotron Rd, Berkeley, CA 94720, USA
Fax: 510 486 6798
Email: http://www.fruitfly.org/EST_estefruitfly.berkeley.edu
hit genomic sequence AE003472; hit genomic sequence AE003525
Plate: AT.246 row: C column: 5
High quality sequence stop: 585.
Location/Qualifiers

FEATURES

source

1. 652
/organism="Drosophila melanogaster"
/db_xref="taxon:7227"
/clone_lib="AT24629"
/sex="male"
/dev_stage="0-3 day old Ore-R males"
/lab_host="DH5-alpha or DH5-alpha Tona as per database (AT
121 on are in Tona cells)"
/note="Organ: Adult testes; Vector: pOTB7; Site_1: EcoRI;
Site_2: XhoI; The mRNA for the testis library was made
from testes and seminal vesicles hand dissected from 0-3
day old Ore-R males. RNA kindly provided by the lab of
Margaret Fuller. Sized fractionated cDNAs were directly
ligated into pOTB7. Plasmid cDNA library."
BASE COUNT 134 a 210 c 184 g 124 t
ORIGIN

alignment_scores:
Quality: 91.00 Length: 189

Ratio: 0.989 Gaps: 7
Percent Similarity: 48.677 Percent Identity: 23.280

alignment_block:
US-09-462-846-2 x BF488952 ..

Align seg 1/1 to: BF488952 from: 1 to: 652

```

155 ArgValValysValysPro..... 160
156  |||||..... 161
90  CGCCCGAAGACGTGCCCCACTGCGATCAGCAGGACACCCCTGAGACAT 139
161  .GlyAspPhePheTyrValProSerGlyThrValHisAlaIle.....G 175
140  TGGCAACAGTGTTCGACCTCTCCGCCCGGAGCAAGATCGAGACAC 189
175  lYlYsGlylIeLeuAlaLeuGlutThrGlnGlnAsnSerAspThrThrTyr 191
190  ACAAGGCGATCTTCGCATCTGCGACCGCAGCAGCAGATGTCGACGAC 239
192  ArgLeuTyrAspTyrAspArgLysAspAla..... 201
240  AAGTGGGGGACTGTCGCTGGCGAGCGCTTCGCTCGTGGCTGCTCT 289
202  ...GluGlyLysLeuArgGluLeuHisLeuLysSerLysSerLysGluValI 217
290  TAAGAGAGGC.....ATCCACGTTCGACTATCCGCGCGAGTGG 327
217  lEglValProSerLLeProGluArgHisThrValHisHisGluGluIle 233
328  TGGACCGCGCGACCTTTCGACCGCGACCGACCTCTCGTCGACACACAGTCC 377
234  GluAspLeuLeuThrThrThrLeuIleGluLysAlaTyr..... 246
378  GAGGACAGAGGTGTGTACACTGACAGCAGCAGCTTATCCCGACACAGC 427
247  ...PheSerValGlyLysTrpAsnLeuSerGlySerAlaSerLeuLysG 262
428  GCCCTACTCTCTTCACACAGCTCCCTGTGCGAGTGGCGCTGTGCT 477
262  lInGlnLysProPheLeuLeuLeuSer.....ValIleGluGlu 275
478  TCGACACAGCGCTACTCCATGCGCAGTCCAGATGCTGTGTCATGTGGGAA 527
276  GlYArG..... 277
528  GCCAGTTTGGGACTTCTGCACACGCGCGCAGTGCATGACAGCTT 577
278  .MetIleSerGlyLysValTyrValTyrProPheLysLysGlyAspHisMetI 294
578  CATACCCAGCGCGCGAGCAAGTGGGTGGCCAGTCCGGAGTCCGTCAATGT 627
294  euLeuProTyrGlyLeu 299
628  TGTCTGCCACACAGCATG 644

seq_name: gb_est90:BF638871

seq_documentation_block:
LOCUS      BF638871          663 bp      mRNA          EST
DEFINITION Nf079F08P1F1073 Phosphate starved leaf Medicago truncatula cDNA
ACCESSION  BF638871
VERSION    BF638871.1 GI:11903029
KEYWORDS   EST.
SOURCE     Medicago truncatula
ORGANISM   Medicago truncatula
            Eukaryota; Viridiplantae; Embryophyta; Tracheophyta; Spermatophyta;
            Magnoliophyta; eudicotyledons; core eudicots; Rosidae; eurosids I;
            Fabales; Fabaceae; Papilionoideae; Medicago.
REFERENCE  Liu,J., Scott,A.D., Harris,A.R., Gonzales,R.A., Bell,C.J., Flores
AUTHORS    H.R., Inman,J.T., Weller,J.W., May,G.D. and Harrison,M.J.

```

TITLE Expressed Sequence Tags from the Samuel Roberts Noble Foundation
JOURNAL Medicago truncatula phosphate-starved leaf library
COMMENT Unpublished (2000)
CONTACT Harrison MJ
PLANT Plant Biology Division
FUNDATION The Samuel Roberts Noble Foundation
ADDRESS 2510 Sam Noble Parkway, Ardmore, OK 73402, USA
TEL Tel: 580 221 7325
FAX Fax: 580 221 7380
EMAIL Email: mjharrison@noble.org
INSERT Insert length: 663 Std Error: 0.00
PLATE Plate: 079 Row: F Column: 08
SEQPRIMER Seq primer: TCACACGAGAACGCTATGAC.
LOCATION Location/Qualifiers

FEATURES
source
1. 663
/organism="Medicago truncatula"
/db_xref="taxon:3880"
/clone="NF079F08PL"
/clone_lib="phosphate starved leaf"
/issue_type="leaf"
/dev_stage="trifoliolate"
/notes="Vector: lambda Zap. At the trifoliolate stage, M.
truncatula plants were transplanted to phosphate-free sand
and grown for a further 30 days. During this 30 day
period, the plants were fertilized twice weekly with 1/2
Hoslands solution containing only 200µM potassium
phosphate. RNA was prepared from above ground tissues."
BASE COUNT 217 a 107 c 157 g 180 t 2 others
ORIGIN

alignment_scores:
Quality: 90.00 Length: 204
Ratio: 0.841 Gaps: 10
Percent Similarity: 52.451 Percent Identity: 23.039

alignment_block:
US-09-462-846-2 x BF638871 ..

Align seg 1/1 to: BF638871 from: 1 to: 663

```

103  AlaAsnIleHisGluAsnGlyLeu.....GlyLysThrG1 115
    |||||..... 116
26  TCAATCTACATACACACATGAGATGAGATGACACAGCAGTGTGCAACAA 75
115  uCyStrpTyrIleIleAspCysGlnLysAspAlaGluIleIleTyrGlyH 132
    |||||..... 133
76  CAAGCATACACATGAT...GAGAGTATGAGAGAGCATATTGTGGCC 122
132  .IsaAsnAla.....ThrThrLysGluLeuThrThrMetIleGlu 145
123  TTAATAGAGGAAGTTCTTACATGCGAGATCTTAAGATATATGCCCA 172
146  .....ArgGlyGluTrpAspGluLeuLeuArgTyrValys..... 157
173  AATTTTGGGAAAGACCAACCAAGGTTACTTAATGGCTTAATAGTTT 222
158  .ValLysProGlyAspPhePheTyrVal.....ProSerGlyThrValH 172
223  TGTGAGCCTGTGATACATGCGCATATATGCGTCTTGTGCTGCGAA 272
172  .IsaAlaIleGlyLysGlylIeLeuAlaLeuGluThrGlnGlnAsnSerAsp 188
273  AATCCACTTGTGCTGATACACTTGCAGGTACACTGCAAAAAAATGTGGTA 322
189  ThrThrTyrArgLeuTyrAspTyrAspArgLysAspAlaGluGly..... 203
323  ATACAGGAATGTTTCTCTCATAGGAGAAAAAAACCTCAGCAGTATGG 372
204  .....LysLeuArgGluLeuHisL 210
373  TTTGTGTGATACGTACACAGAGATGTGTGTATAGCAACTTTCACAG 422

```

```

210 eulyslysSerIleGluVal.....IleGluValProSerIlePro 223
      |||||.....
423 TGAAGAAACAAATACATATTCAGCACATTTCAGTTCATCA..... 466
224 GUAUgHsThrValHsHsIsgIuInIleGluAspIleuThrThTh 240
      |||||.....
467 .....ACAATGTCCAAAGAGAAATCAACGGCTTAGTTGATGCAC 507
240 rLeuIleGlu.....CysAlaTyrPheSerValGlyLysT 252
      |||||.....
508 AATCATTAAGATGGCTCCATGATGATGAGTAAATAGGTAAT 557
252 rPasnLeuSerGlySerAlaSerLeuIleGluInIleLysProPheLeu 268
      |||||.....
558 GGCATTGTGANAAGTATAGTGGTGGCGAAGAAAGAAACTGATTC 607
269 IleSerValIle 272
      |||||.....
608 CTGAGATTCCTC 619

seq_name: gb_est71:BE249341

seq_documentation block:
LOCUS BE249341 531 bp mRNA EST 21-DEC-2000
DEFINITION NF014C07LF1053 Developing leaf Medicago truncatula cDNA clone
ACCESSION NF014C07LF 5, mRNA sequence.
VERSION BE249341.2 GI:11959569
KEYWORDS EST.
SOURCE barrel medic.
ORGANISM Medicago truncatula
Eukaryota; Viridiplantae; Embryophyta; Tracheophyta; Spermatophyta;
Fabales; Fabaceae; Papilionoideae; Medicago.
1 (bases 1 to 531)
Torres-Jerez,I., Scott,A.D., Harris,A.R., Gonzales,R.A., Bell,C.J.,
Flores,H.R., Inman,J.T., Weller,J.W. and May,G.D.
Expressed Sequence Tags from the Samuel Roberts Noble Foundation
Medicago truncatula leaf library
Unpublished (2000)
On Jul 13, 2000 this sequence version replaced gi:219401.
COMMENT
Contact: May GD
Plant Biology Division
The Samuel Roberts Noble Foundation
2310 Sam Noble Parkway, Ardmore, OK 73402, USA
Tel: 580 221 7391
Fax: 580 221 7380
Email: gdmay@noble.org
Medicago Genome Initiative accession: MGI:S:14889
Insert length: 659 Std Error: 0.00
Plate: 014 row: C column: 07
Seq primer: TCACACAGAAACGCTATGAC.

FEATURES
Source
1..531
/organism="Medicago truncatula"
/db_xref="taxon:3880"
/clone="NF014C07LF"
/clone_lib="Developing leaf"
/tissue_type="leaf"
/dev_stage="Pooled developmental"
/notes="vector: lambda zap; Contains a mixture of very
young, developing, mature and senescing leaves."

BASE COUNT 165 a 95 c 134 g 137 t
ORIGIN
alignment_scores:
Quality: 89.00 Length: 146
Ratio: 1.290 Gaps: 6
Percent Similarity: 47.260 Percent Identity: 25.342

alignment_block:
US-09-462-846-2 x BE249341 ..

```

```

Align seg 1/1 to: BE249341 from: 1 to: 531
80 ProLeuLeuThrLysIleLeuAspAlaAspGlnAspLeuSerValGlnVal 96
      |||.....
4 CCTCTTCGCAATCAAAACTTGCCTAAACCCAGACGATCCTCCTCATCTT 53
96 LHisProAsnAspGluTyrAlaAsnIleHsIsgLusnGlyGlu..... 110
54 TCAT...AACACCGTCGATAGAAATGATGATGAGAGAGGTAAGGCTGAC 100
110 .....
101 AGTGAGTTGAAGTGTATCATGAGGAGCCTAGAGCTTTGTATCATCAAT 150
111 ..LeuGlySerThrGlyCysTrpTyrIleIleAspGlyLys..... 124
151 TTCTGACCAAGAGAGATGTGATGTCGACATAGCAAAAGCCGAG 200
125 .....AspAlaGluIleIleTyrGlnHisAsnA 134
201 TATGCACAGTCAACGGTGTGTATAGCGAGACTGGAAAGCAAAAGACA 250
134 IaThrThrLysGluGluLeuThrThrMetIleGluArgGlyGluTrpAsp 150
251 GCAGAGTGTCTACTAGCTCTGGAACATTCTGCGCCAGGGAGAGT...GAT 297
151 GluLeuLeuArgArgValLysValLysProGlyAspPhePheTyrValPr 167
298 AAGATTGTGAGAGATATTGAGAAAAAATTCGCCACTTACTTTATTTC 347
167 cSerGlyThrValHisAlaIleGlyLysGlyIleLeuAlaGluThrng 184
348 T.....GTAGACGACCGTGAAGACTTCACTGACTTATGAAAGTTG 391
184 InGlnAsnSerAspThrThrTyrArgLeuTyrAspTyr 196
392 GACAAAGATATGAGCTCTAC.....TATGACTAC 420

seq_name: gb_est88:BF493329

seq_documentation block:
LOCUS BF493329 712 bp mRNA EST 06-DEC-2000
DEFINITION AT01657.5prime AT Drosophila melanogaster adult testes pothB7
ACCESSION BF493329
VERSION BF493329.1 GI:11576630
KEYWORDS EST.
SOURCE fruit fly.
ORGANISM Drosophila melanogaster
Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
Pterygota; Neoptera; Endopterygota; Diptera; Brachycera;
Muscomorpha; Ephydroidea; Drosophilidae; Drosophila.
1 (bases 1 to 712)
Stapleton,M., Brkstein,P., Hong,L., Agbayani,A., Baxter,E., Berman
,B., Carlson,D., Champe,M., Chavez,C., Chavet,M., Dossett,V., Farfan
,D., Frise,E., George,R., Gonzalez,M., Guarin,H., Harris,N., Li,P.,
Liao,G., Miranda,A., Misra,S., Mungall,C.D., Nunoo,J., Pacleb,J.,
Park,S., Paragas,V., Phouaneavong,S., Wan,K., Yu,C., Celiker,S.,
Lewis,S.E. and Rubin,G.M.
Berkeley Drosophila Gene Collection Project
Unpublished (2000)
Contact: Stapleton, M.
BDGP
Lawrence Berkeley National Lab
One Cyclotron Rd, Berkeley, CA 94720, USA
Fax: 510 486 6798
Email: http://www.fruitfly.org/EST, est@fruitfly.berkeley.edu
hit genomic sequence AB003472; hit genomic sequence AB003525
Plate: AT.16 row: E column: 9
High quality sequence stop: 479.
Location/Qualifiers
FEATURES

```

```
source 1.712
/organism="Drosophila melanogaster"
/db_xref="taxon:7227"
/clone_lib="AT01657"
/sex="male"
/dev_stage="0-3 day old Ore-R males"
/lab_host="DHS-alpha or DHS-alpha Tona as per database (AT
121 on are in Tona cells)"
/note="Organ: ADULT testes; Vector: pOTB7; Site_1: EcoRI;
Site_2: XhoI; The mRNA for the testis library was made
from testes and seminal vesicles hand dissected from 0-3
day old Ore-R males. RNA kindly provided by the lab of
Margaret Fuller. Sized fractionated cDNAs were directly
ligated into pOTB7. Plasmid cDNA library."

BASE COUNT 143 a 228 c 204 g 137 t
ORIGIN

alignment_scores:
Quality: 88.50 Length: 202
Ratio: 0.932 Gaps: 8
Percent Similarity: 47.030 Percent Identity: 23.762

alignment_block:
US-09-462-846-2 x BF493329 ..

Align seg 1/1 to: BF493329 from: 1 to: 712

155 ArgValLysValLysPro..... 160
|||||:|||||
86 CCGCTGAACACTGTCGCCCATGCGATGACGACGACCCCTGAAGACGAT 135
|||||:|||||
161 GLVSPHPHETRYVALPROSERGLYTHVALHSAIAIE.....G 175
|||||:|||||
136 TGGCAACATGTTTCGACTCTCCGCCCGCGAGACACAGTTCGAGACAC 185
|||||:|||||
175 LYSGLYIIEUALALEUGLUTHRGLNINASNSERSAPYTHRYT 191
|||||:|||||
186 MCNAGGCGATCTTCGCNTACTGCGCCGCGCAGCAGATGTCGACGAC 235
|||||:|||||
192 ARGLEUTYRASPTRYASPARGLYSASPALA..... 201
|||||:|||||
236 AAGTGGGGGACTGTCGCTGGGGAGCGCTTCGCTCGTCCGCTCCT 285
|||||:|||||
202 ...GLUGLYLSLEUARGGLUENHISLEULYSERLLEGUVALI 217
|||||:|||||
286 TAAGGAGGGC.....ATCCACGTTCGACTATCCGGCCAGGATG 323
|||||:|||||
217 LEGUVALPROSERLEPROGLUARHISTHRYVALHISHISGLUINILE 233
|||||:|||||
324 TGGACGCGGCGACCTTTCGACCGGCGACCGCTCCTGCACCCAGTCC 373
|||||:|||||
234 GUASPPLLEUTHRTHRTHRLEULIEGLUCYALATYR..... 246
|||||:|||||
374 GAGGCAAGGTGTGTACACTGACGTGACACCTTATCCCGACGAGGC 423
|||||:|||||
247 ...PHESEVALGLYSTRPASNLEUSERGLYSERLASERLAULYS 262
|||||:|||||
424 GCCCTACTCCGTCYCAACAGCTCCCTGCGAGTGTGGCTTGGCT 473
|||||:|||||
262 INGLILYPROPHLEULEULESER.....VALILEGLUGLIGLU 275
|||||:|||||
474 TCGACACTGCTACTCCATGCGCCAGTCCCAATGCTGTGTCATGTGGGA 523
|||||:|||||
276 GLYATG..... 277
|||||:|||||
524 GGCCAGTTTGGGACTTCTGCACACGCGCGCAGTGCATCGACACGTT 573
|||||:|||||
278 METILESERGLYGLUTRYVALTYRPROPHETLSYGLYASPHISMETL 294
|||||:|||||
574 CATACCCAGCGCGGAGACAGTGGTGGCCGCGCATCCGAGTGCATCT 623
|||||:|||||
```

```
294 euleuprotyr.....GlyLeuGlyGluPhelLysLeuGluGlyTyr 307
|||||:|||||
624 TGCTGCCACACAGCATGAGAGCGCATGCGCCGAGACACTCTTCCGCGCGC 673
|||||:|||||
308 AAGIu 309
|||||
674 ATGAG 679
```

Fri May 18 10:38:05 2001

us-09-462-846-2.p2n.rst

Page 13

•
•
•
•


```

518 GCAGCAGCCCATGCTGACTCTG...ATGGCGCGCATCCCAAGCA 564
47 lnservValValGlnAsnGlyMetTyrGlySerGlyPheThrLeuSerGlyLeu 63
   |||:::||||:::
565 GCTCCGGAATCACCCACCGCCGCAACCGCTCTCTCCCTGCGTGCAGCC 614
   |||:::||||:::
64 TrpGlnHisLysArg...HisLeuPheGlyGlnLeuGlyLysAspArgPhe 79
   |||:::||||:::
615 ATCGAAGAAAGATAAACCGCATGCTGGCGAAGCGGTAGCCACCGCTT 664
   |||:::||||:::
79 e.....ProLeuLeuThrLysIleLeuAspAlaAspGlnAspLeu 93
   |||:::||||:::
665 CGCGCAACTGCGCTTCTGTTTAAAGTACTGCGCGCCCAACCGCTCT 714
   |||:::||||:::
715 CTATTCAGGTGACCCGCAAT.....AAACGCACTCC 746
   |||:::||||:::
110 GluLeuGly..... 112
747 GAAATCGGTTTGGCAAAAGAAATGCGCGGATCCCATCCCATGATCCGC 796
   |||:::||||:::
113 .....LysThrGlyCysTyrTyrIleI 120
   |||:::||||:::
797 AGAGCGGAACATTAAAGATCTTAACCATTAACGAGACTGTTTGGCC 846
   |||:::||||:::
120 leasp.....CysGlnLysAspAlaGluIleI 129
   |||:::||||:::
847 TGACCGCTTCTCTGCGATGAACCGCTTCCCGCAATTTTCTGACATTGC 896
   |||:::||||:::
130 Tyr.....GlyHisAsnAla..... 134
   |||:::||||:::
897 TCTTACTGCACTGTCCGCGCGCATTCGCTATCCGCTATCCCATTTT 946
   |||:::||||:::
135 .....ThrThrLysGlnGluLeuThrPheMetIle..... 144
   |||:::||||:::
947 GCAGGCGCGAATGCTGAACGCTGACGACGACTTTTCCGACGCTTGA 996
   |||:::||||:::
144 ..... 144
997 AATATGCAAGCGCAAGAAATCCGCGCGCTTATGCGCTACTCAAAAGCGCG 1046
   |||:::||||:::
145 .....GluArgGlyGlu...TyrAspGlu..... 151
   |||:::||||:::
1047 CTTAAGACGACAGAGCGGCAACCTGGCAAAAGATCCGCGTATTCAGA 1096
   |||:::||||:::
152 .....LeuLeuArgArgY 156
   |||:::||||:::
1097 GATATATCTCTGACGACAGCGGCTTTCTCTCTCTTGTGCTGCAATGTGG 1146
   |||:::||||:::
156 allYsValLysProGlyAspPhePheTyrValProSerGlyThrValHis 172
   |||:::||||:::
1147 TCAAACTGAATCCGCGAGCGGATGTTCTGTTGCTGAAAGCGCTCAT 1196
   |||:::||||:::
173 AlaIleGlyLysGlyIleLeuAlaLeuGluThrGlnGlnAsnSerPhe 189
   |||:::||||:::
1197 GCTATCTGCAAGGCGCTT...GCCGTGAAGTATGCGCACTCCGATAA 1243
   |||:::||||:::
189 rThrTyrArgLeu..... 193
   |||:::||||:::
1244 CGTTCGCGCGCTGCTTACGCCAAATATATGACATCCCTGAGCTGG 1293
   |||:::||||:::
194 .....TyrAspTyrAspArgLysAspAlaGluGlyLysLeuArgGlu 207
   |||:::||||:::
1294 TGGCGAAGCTGAAGTTGCAACCTTAAGCGCTGC...GGCGGTTGCTGACT 1340
   |||:::||||:::
208 LeuHisLeuLysLysSerIleGlnVal...IleGlnValProSerIlePhe 223
   |||:::||||:::
1341 GCCCGGTGAAGAAAGCGCGCGAGCTGAGCTCCCAATTCGGGTGACGA 1390
   |||:::||||:::
223 oGluArgHisThrValHisHisGlnGlnIleGluAspLeuLeuThrThT 240
   |||:::||||:::
1391 TTTTGCTTTTTCATGCAAGCACTGGCGCTTCCAGAGACG..... 1430

```

```

240 hLeuIleGlyCysAlaTyrPheSerValGlyLysTyrAsn..... 253
1431 .....AGCATCGCGCCAAACAGACGCCCGGATT 1457
254 .....LeuSerGlySerAlaSerLeuLysGlnGlnLysProPhe 267
   |||:::||||:::
1458 CTGTTCTGCGTTGAGCGGAGCGCGTGTACTGTAAGATGAACAGCGCT 1507
   |||:::||||:::
267 uLeuLeuSerValIleGlyGlyGlyArgMetIleSerGlyGlu 282
   |||:::||||:::
1508 GGTACTGAG.....CCGGTGAATCTGCTTATTCGCGCGGAT 1547
seq_name: /cgn2.6/prodata/1/lna/6B.COMB.seq:US-09-001-982-9
seq_documentation_block:
; Sequence 9, Application US/09001982
; Patent No. 6204246
; GENERAL INFORMATION:
; APPLICANT: Bosch, Hendrick J.
; APPLICANT: Stiekema, Willem J.
; TITLE OF INVENTION: Hybrid Toxin
; NUMBER OF SEQUENCES: 15
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: No. 6204246artis Corporation
; STREET: 3054 Cornwallis Road
; CITY: Research Triangle Park
; STATE: NC
; COUNTRY: USA
; ZIP: 27709
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/001,982
; FILING DATE:
; CLASSIFICATION:
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/602,737
; FILING DATE: 21-FEB-1996
; ATTORNEY/AGENT INFORMATION:
; NAME: Meigs, J. Timothy
; REGISTRATION NUMBER: 38,241
; REFERENCE/DOCKET NUMBER: 130-4080/PCF/CIP
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 919-541-8687
; TELEFAX: 919-541-8689
; INFORMATION FOR SEQ ID NO: 9:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 3468 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: cDNA
; ORIGINAL SOURCE:
; ORGANISM: Bacillus thuringiensis
; FEATURE:
; NAME/KEY: CDS
; LOCATION: 1..3468
; US-09-001-982-9
alignment_scores:
Quality: 92.50 Length: 338
Ratio: 0.571 Gaps: 18
Percent Similarity: 47.929 Percent Identity: 18.639
alignment_block:
US-09-462-846-2 x US-09-001-982-9
Align seg 1/1 to: US-09-001-982-9 from: 1 to: 3468

```

```

24 AspphegLYTThrThrlleProserGlnArgThrglyGluCysTrpAlaph 40
|||||.....:.....:.....:.....:.....:.....:
2203 GATTTAAATCAATCAATAGTACAGAAATAATGCA.....TGGAAAGC 2246
40 eAlaAlaHISGlnAsnGlyInserValGlnAsnGly...MetTYR 56
:::|||||.....:.....:.....:.....:.....:.....:
2247 AGT.....AACGGCTACTATTAGTGAAGCGCGGCATTCTATA 2287
56 ysgLY.....PheThrLeuSerGluLeuTrpGluHISHis...ArgHIS 69
|||||.....:.....:.....:.....:.....:.....:
2288 AAGGCGGTGCAATCAGCTAGCAAGTGCACGAAATAATACCAACATAC 2337
70 LeuPheGlyGlnLeuGluGlyAspArgPheProLeuLeuThrLys..... 84
:::.....:.....:.....:.....:.....:.....:
2338 ATCTATCAAAAAGTAGATGATCGAGTTAAAGCCGTATACGTTATAG 2387
85 .....IleLeuAspAlaAspGlnAspLeuSerValGlnVal.... 96
:::.....:.....:.....:.....:.....:.....:
2388 ACTGATGGGTCGTGTAAGAGTAGTCAAGATTAGAAATGATCTCATTC 2437
96 .....:.....:.....:.....:.....:.....:
2438 ACCATCATTAAGTCATCTTGAAAAAATGTACCAAGATAATTAGTATCT 2487
97 .....HisProAsnAspGluTyrAlaAsnIleHISGlu..... 107
:::|||||.....:.....:.....:.....:.....:.....:
2488 GATACTTACCAGATGATCTGTAGTGAATCAATCAATCGTCAAGACA 2537
108 .....AsnGlyGluLeuGlyLysThrGluCysTrpTyrIleIleA 121
|||||.....:.....:.....:.....:.....:.....:
2538 ACAGATGTAATAGCGCAACTG...GAACACAGACATCATCATCCGAGG 2584
121 sPcysGlnIysAspAlaGluIleIleTyrGlyHISAsnAlaThrThrLys 137
|||||.....:.....:.....:.....:.....:.....:
2585 ATTGCTGTAGAGAGCTCAACACATGAGTTTCTCCATATATGATACA 2634
138 GlnGluLeuThrThrMetIleGluArgGlyLysTrpAspGluLeuLar 154
:::|||||.....:.....:.....:.....:.....:.....:
2635 GGGGATTTAAATTCGAGTGTAGACCGGAATCTG.....GCGATCTT 2678
154 garValIysValIysProGlyAspPhePheTyrValProserGlyThrV 171
:::|||||.....:.....:.....:.....:.....:.....:
2679 TAAAGTTCAGACACCGATGT.....TATGCGAGCTAGGAATTC 2719
171 aHISAlaIleGlyLysGlyIleLeu.....AlaLeuGluThrGln 184
:::|||||.....:.....:.....:.....:.....:.....:
2720 TTGAATTGTGAGAGTGGACCGCTATCGGTGAATCTTTAGAACGTGA 2769
185 GlnAsnSerAspThr.....:.....:.....:.....:.....: 189
|||||.....:.....:.....:.....:.....:.....:
2770 CAAGGATATATCAAAATGAGTGCAGAGTACGAGAAACCGTGCAGA 2819
190 ThrTyrArgLeuTyr.....:.....:.....:.....:.....: 194
|||||.....:.....:.....:.....:.....:.....:
2820 AACAGATCGGCTGTATCAGATGCCAAACATCATCATTTATTG 2869
195 ..AspTyr.....:.....:.....:.....:.....:.....: 198
|||||.....:.....:.....:.....:.....:.....:
2870 TGGATTATCAAGATCAACATTAATCCAGAAATAGGATGCGAGATATT 2919
199 LysAspAlaGlnGluGlyLysLeuArgGluLeuHISLeuLysSerIleG 215
|||||.....:.....:.....:.....:.....:.....:
2920 ATGAGAGCTCAAAATCTGTGCGATCATTTGAGATATAGCGAGTGC 2969
215 uValIleGluValProserIleProGluArgHISThrValHISHisGlu 232
|||||.....:.....:.....:.....:.....:.....:
2970 CGTACTGCAAAATCCCTGGAATT.....AACTATG 2998
232 InIleGluLysLeuLeuThrThrlleGluCysAlaTyrPhe... 247
:::|||||.....:.....:.....:.....:.....:.....:
2999 AGATTATACAGAGAGCTGCCAATCGCTTACAAACAGCATCGATCTGAT 3048

```

```

248 .....SerValGly 251
3049 AGCTCTCGAATGCGGTGCAAAATGGGACCTTTAAACAGGGGTAGATAG 3098
251 sTrpAsnLeuSerGlySerAlaSerLeuLysGln.....GlnLysPro 266
|||||.....:.....:.....:.....:.....:.....:
3099 CTGGAATGCAACAGCGGGGTGATCGGTACAAACAGATGCAATACGATT 3148
266 heLeuLeuIleSer 270
|||||.....:.....:.....:.....:.....:.....:
3149 TCTTACTTCTTCT 3162

seq_name: /cgn2_6/ptodata/1/ina/5A_COMB.seq:us-08-156-866-1

seq documentation block:
Sequence 1, Application US/08156866
Patent No. 5443831
GENERAL INFORMATION:
APPLICANT: Keeler, Jr., Calvin L.
APPLICANT: Poulsen, David J.
TITLE OF INVENTION: Gene Encoding Glycoprotein B of
TITLE OF INVENTION: Infectious Laryngotracheitis Virus
NUMBER OF SEQUENCES: 1
CORRESPONDENCE ADDRESS:
ADDRESSEE: Connolly and Hutz
STREET: 1220 Market Building
CITY: Wilmington
STATE: Delaware
COUNTRY: U.S.A.
ZIP: 19899
COMPUTER READABLE FORM:
MEDIUM TYPE: 3.5 inch diskette
COMPUTER: IBM/PC or Compatible
OPERATING SYSTEM: MS-DOS
SOFTWARE: WordPerfect 5.1
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/156,866
FILING DATE: No. 5443831 yet assigned
CLASSIFICATION: 424
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US/07/788,123
FILING DATE:
INFORMATION FOR SEQ ID NO: 1:
SEQUENCE CHARACTERISTICS:
LENGTH: 3065 base pairs
TYPE: Nucleic Acid
STRANDEDNESS: double stranded
TOPOLOGY: linear
MOLECULE TYPE: Genomic DNA
HYPOTHETICAL: NO
ANTI-SENSE: NO
ORIGINAL SOURCE:
ORGANISM: Gallid Herpesvirus 1
INDIVIDUAL ISOLATE: 632
US-08-156-866-1

alignment_scores:
Quality: 89.50 Length: 355
Ratio: 0.581 Gaps: 16
Percent Similarity: 43.380 Percent Identity: 19.437

alignment_block:
US-09-462-846-2 x US-08-156-866-1 ..

Align seg 1/1 to: US-08-156-866-1 from: 1 to: 3065

1 MetThrThrglnProLeuPhePheLysProValPheLysGluArgIleTr 17
:::|||||.....:.....:.....:.....:.....:.....:
548 GTTACTACAGCGGGCATTAATTCACAGCCCAAAATACAAATGATACGT 597
:::|||||.....:.....:.....:.....:.....:.....:
17 pGlyGlyThrAlaLeuAlaAspPheGlyTyrThrIleProserGlnArgT 34
:::|||||.....:.....:.....:.....:.....:.....:

```

```

598 GACGAGGGTCCAAATA...GACTATCATGAAATGTCAGGATGATCAT 644
34 hcglyglucystrpalaphlealialahslnglyglinservall 50
   ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
645 CGGAGAAATGCTCATCCAAAGCAACGTAT..... 673
51 gluasnglymettyrlysglypherthleusergluleutrglnhishl 67
   ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
674 .....CAVAAAAATTCATGTTTTCGAGCTTACGACCAATCA 711
67 sarghislsleupheglygluleuglyasparcypheproleuthrl 84
   ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
712 TCAA.....CGAGAAAAAATTCCTCCCTGCTCCAT 743
84 ysileleuaspalaaspglinspleuservalglnvalhisproasasp 100.
   ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
744 CACTGTATAGATCA.....ACTGTCTCAAGCGTTTCATACAACTAAC 787
101 glutyralaasnlehisgluasnglygluleuglyysrthrcglucysr 117
   ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
788 ...TTTACTAAGCCGACATCAACC..... 808
117 ptyrilleleaspcysglinsaspalaglnleiletyrlyhisasna 134
   ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
809 .....CTGGATACCGAA 821
134 latthrthlysgluleuthrthrmethleuglnatrglylutrpsap 150
   ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
822 CGCTTACATCGGTCCACTGTGTGTG..... 847
151 gluleuauarqarqvalysvalysploglasphepherthvalpr 167
   ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
848 GAATATCTACAGCGATCTGTATACCGTATGATTAATCTTGCAATGCG 897
167 osercllythrvalhisalileglylysglyleleualaleuglnthc 184
   ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
898 GACAGGTATACAGTAGAATTTCTCTTTTACCAAAAACAGCAGCG 947
184 lnglnasnsersapthrtthrtlyarqleuayr..... 194
   ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
948 GACCAAGCGCTCACAGTCTACAGACATATGATTTTCGAATGCGA 997
195 aspyrtaaparglyasapalaglu...glylyleuauarglu..... 207
   ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
998 AATATATCAAGTACAGGATTGGAACCGCAAAATAGACCCCTTAAAA 1047
207 ..... 207
1048 AAGAAACTTCTAACAGATGAACAATCAGTATAGCGTGGATGCAATGG 1097
208 .....leuhlsleuylslyserillegluvalileglu 218
   ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
1098 AAGAAAGCAATCTGTATGTACTCTCAGTAAATGATGAAAGTCCCGGAA 1147
219 valproserilleprogluarghsthvalhisglucnleuglnas 235
   ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
1148 GCAGTTCGTTCGTACAAAAACAGT...TACCACTTTTCACTTAAGA 1194
235 pleuueuthrthrlleuileglucysalatyrpheservaleglyls. 251
   ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
1195 TATGACTATGACG.....TTCTGCTCCGGAAGAAC 1223
252 .....trpasnleuserglyseralaserleuysglnglninsprophe 266
   ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
1224 AACCTTTTAAATCAGC...AGCTCATTTGGCTGAATGCGTCTTACC 1270
267 leuleuileservalileglugly..... 274
   ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
1271 ATAGCCACGAGGCAATGATGATGATCTTTCAGAAAGTATAGTTCGAC 1320
275 .....gluglyargmettleserglyglutryv 284
   ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
1321 TCATGTCCGTCTGTGGGACATCGAAATACATCTCGAGTGGGCGATTTTC 1370

```

```

284 altyrprophelyslysglyasphismetleuauprotyrlyleugly 300
   ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
1371 TCATGCCATTTCAGAAA.....CTCATGACCATGCTTGCTTGGCT 1408
301 gluphelysleuglu 305
   ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
1409 GAATGTACTAGAA 1423

```

seq_name: /cogn2_6/ptodata/1/ina/5b_comb.seq:us-08-813-940-24

seq_documentation_block:

```

; Sequence 24, Application US/08813940
; Patent No. 5834279
; GENERAL INFORMATION:
; APPLICANT: Rubin, Harvey
; APPLICANT: Yang, Fude
; APPLICANT: Avarbock, David
; APPLICANT: Curran, Sean
; TITLE OF INVENTION: Methods of Identifying Compounds that
; TITLE OF INVENTION: Inhibit DNA Synthesis in Mycobacterium Tuberculosis and
; TITLE OF INVENTION: Composition, Reagents and Kits for Performing the Same
; NUMBER OF SEQUENCES: 25
; CORRESPONDENCE ADDRESS:
; ADDRESS: Woodcock Washburn Kurtz Mackiewicz & No. 5834279-15
; STREET: One Liberty Place, 46th Floor
; CITY: Philadelphia
; STATE: Pennsylvania
; COUNTRY: USA
; ZIP: 19103
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: Windows
; SOFTWARE: Wordperfect
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/813,940
; FILING DATE:
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/612,271
; FILING DATE: 07-MAR-1996
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: Deluca, Mark
; REGISTRATION NUMBER: 33,229
; REFERENCE/DOCKET NUMBER: UPAP-0220
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 215-568-3100
; TELEFAX: 215-568-3439
; INFORMATION FOR SEQ ID NO: 24:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 4107 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: double
; TOPOLOGY: both
; MOLECULE TYPE: cDNA
; FEATURE:
; NAME/KEY: CDS
; LOCATION: 446..2620
; US-08-813-940-24

```

alignment_scores:

```

Quality: 84.00 Length: 276
Ratio: 0.672 Gaps: 12
Percent Similarity: 45.280 Percent Identity: 22.826

```

alignment_block:

US-09-462-846-2 x US-08-813-940-24 ..

Align seg 1/1 to: US-08-813-940-24 from: 1 to: 4107

51. GlnAsnGlyMetYrIyrsGlyPheThrLeuSerGluLeuTrpGluHisIle 67
1217 CACCTGGGCGCTCTCAGGAGCGCGCGGCTGTACCTGCACGCCCATCA 1266
67 sacGHisIlePheGlyGlnLeuGluGlyAspArgPheProLeuLeuThrL 84
1267 CCCCCGACATTAC.....CGATTTC...CTGGACACCA 1295
84 ysiIleuAspAlaAspGlnAspLeuSerValGln..... 95
1296 AGCGTGAAGACGCGCAGAGAAAGATCCGGATCAAGACCGCTGAGTCGGGG 1345
96 ...ValHisProAsnAspGluTyrValAlaAsnIleHisGluAsnGlyGlu 111
1346 GTGTGTATCCCGCCGACATCACCTTC.....GAGTT 1374
111 uGlyLysThrGluGlySerTrpTyrIleIleAspGlyGlnLysAspAlaGlu 128
1375 GGGCAAGCGCACACATGACATGATACCTGTCTCGCCCTCAAGATGTGAGAC 1424
128 leileIyryr.....HisAsnAlaThrThrLysGluGluLeuThr 141
1425 GGGGTCTAGGGTGGCGGCTGCTGCTGACATCTCGGTACACGAGAGTACTAC 1474
142 ThreIleIleGluArgGlyGluTrpAspGluLeuLeuArgArgValLysVa 158
1475 GAATATGGTCGAT.....GACGCGCGCATCCGGAAGCCAAAGAT 1512
158 llyAspProGlyAspPhePheTyrValProSerGlyThrValHisAlaIleG 175
1513 CAAGCAGCGGAGATCTTCCACAGCCGTCGCGAGCTGCAGATTCGAGCTCG 1562
175 llyLysGlyIleLeuAlaLeuGluThrGlnGlnAsnSerAspThrThrTyr 191
1563 GCTACCCCTAATATCATGTGTGAA.....GACACCGTCAAT 1597
192 ArgLeuTyrAspTyrAspArgLysAspAlaGluGlyLysLeuArgGlu 208
1598 CGCGCTAATCCAAATTGAT.....GGCAAGATCACGCGACAG 1632
208 uHisLeuLysLysSerIleGluValIleGluValProSerIleProGluA 225
1633 CAACCTGTGCTCGGACATCTCGCAATGTCTACGCCGTATTTGTTCAAC. 1681
225 rGHisThrValHisIleGluGlnIleGluAspLeuLeuThrThrLLeu 241
1682GAGGACTTG..... 1690
242 lIleGluCysAlaTyrPheSerValGlyLys.....TrpAsnLeu 255
1691TCGTATCCCAAAAGGGCGCAAGACATTTCTGCACACCTGGG 1731
255 rGlySerAlaSerLeuLysGlnGlnLysProPheLeuLeuIleSerValI 272
1732 GTTCCTGAACATCGCCAGACATGGAGCTCGCCGGAATTCGGCGAGACGA 1781
272 lIeGluGlyGluArgMetIleSerGly..... 281
1782 TCGAGGTGGGATCGCGCGCTGACCGCGGTGAGGCACCAAAACCATATTC 1831
282 GluTyrValTyrProPheLysGlyAspHisMetLeuLeuProTyrGly 298
1832 AGTCTGGCGCTCAATCGACAGGGGACAAACGATCTCCACGCGATCGG 1881
298 yLeuGlyGluPheLysLeuGluGlyTyr 307
1882 GCTAGACACATGATACCTGCACGGCTAC 1909
seq_name: /cgn2_6/ptodata/1/ina/5B.COMB.seq:US-08-447-464-2
seq_documentation_block:
; Sequence 2, Application US/08447464
; Patent No. 5840842

GENERAL INFORMATION:
APPLICANT: Schlesinger, Joseph
APPLICANT: Yan, Hai
TITLE OF INVENTION: NOVEL RECEPTOR-TYPE PROTEIN
TITLE OF INVENTION: PHOSPHOTYROSINE PHOSPHATASE-STIMULATED
NUMBER OF SEQUENCES: 12
CORRESPONDENCE ADDRESS:
ADDRESSEE: Pennie & Edmonds
STREET: 1155 Avenue of the Americas
CITY: New York
STATE: New York
COUNTRY: U.S.A.
ZIP: 10036-2711
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patent In Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/447,464
FILING DATE: 24-MAY-1995
CLASSIFICATION: 435
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 08/130,570
FILING DATE: 01-OCT-1993
ATTORNEY/AGENT INFORMATION:
NAME: Mistrock, S. Leslie
REGISTRATION NUMBER: 18,872
REFERENCE/DOCKET NUMBER: 7683-043
TELECOMMUNICATION INFORMATION:
TELEPHONE: 212-790-9090
TELEFAX: 212-869-8864/9741
TELEX: 66141 PENNIE
INFORMATION FOR SEQ. ID NO. 2:
SEQUENCE CHARACTERISTICS:
LENGTH: 5690 base pairs
TYPE: nucleic acid
STRANDEDNESS: unknown
TOPOLOGY: unknown
MOLECULE TYPE: DNA (genomic)
FEATURE:
NAME/KEY: CDS
LOCATION: 833..5338
US-08-447-464-2

```

alignment_scores:
  Quality: 84.00      Length: 211
  Ratio: 0.778      Gaps: 11
  Percent Similarity: 51.185      Percent Identity: 21.327

alignment_block:
  US-09-462-846-2 x US-08-447-464-2 ..

Align seq 1/1 to: US-08-447-464-2 from: 1 to: 5690

40 pheAlaAlaHisGlnAsnGlyGlnSerValValGlnAsnGlyMetTyrLy 56
   ::::: ::::: ::::: ::::: ::::: :::::
4091 TTTTCTGTACACAGAAATGGCTCTAGTAGAGAAAGGTGAGTACACATT 4140
   ::::: ::::: ::::: ::::: ::::: :::::
56 sGlyPheThrLeuSerGluTyrPrgLinhHisArgHisLeuPheGlyG 73
   ::::: ::::: ::::: ::::: ::::: :::::
4141 TCACTGTACA.....GCATGGCTGGACCAC.....GGGG 4169
   ::::: ::::: ::::: ::::: ::::: :::::
73 InleuGlnGlyAspArgPheProLeuThrLysIleLeuAspAlaAsp 89
   ::::: ::::: ::::: ::::: ::::: :::::
4170 TACCCGAG.....TACCC.....ACACCTTCTGCGCGTTCTG 4204
   ::::: ::::: ::::: ::::: ::::: :::::
90 GlnAspSerValGlnValHisProAsnAspGluTyrAlaAsnIleHi 106
   ::::: ::::: ::::: ::::: ::::: :::::
4205 CGCAGAGTCAGACCTGCAACCCGCGCTACCGCATGGCCATGTGGTCCA 4254
   ::::: ::::: ::::: ::::: ::::: :::::
106 sGlnAsnGlyLeuGluGlyLysThrGluCysTrrPyrIleIleAspCys 122

```

[illegible]

```

? INFORMATION FOR SEQ ID NO: 2:
? SEQUENCE CHARACTERISTICS:
? LENGTH: 5690 base pairs
? TYPE: nucleic acid
? STRANDEDNESS: unknown
? TOPOLOGY: unknown
? MOLECULE TYPE: DNA (genomic)
? FEATURE:
? NAME/KEY: CDS
? LOCATION: 833..5338
? US-08-716-679-2

alignment_scores:
    Quality: 84.00      Length: 211
    Ratio: 0.778        Gaps: 11
    Percent similarity: 51.185      Percent identity: 21.327

alignment_block:
US-09-462-846-2 x US-08-716-679-2 ..

Align seg 1/1 to: US-08-716-679-2 from: 1 to: 5690

40 Phealaalaahisglnaaglyglnserivaivaiglnasnlymetrily 56
||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
4091 TTTTCTCTACACAGATAGGCTGTAGTACAGAAACCGTACGATTCATT 4140
56 sglypherthireusergluleutrpunhishisatghisleupneglyg 73
||||| ||||| ||||| ||||| ||||| ||||| |||||
4141 TCAGTGTCCA.....GCATGGCCGTGACCC.....GGGG 4165
73 Inleugluclyaspargpheproleuethrlystilleuaspalaasp 89
||||| ||||| ||||| ||||| ||||| ||||| |||||
4170 TACCCGAG.....TACCCC.....ACACCTTCCTGCGCTTTCG 4204
90 GlnaspuuseraValglnvalhisprouasapglutryAlasnlh1 106
||||| ||||| ||||| ||||| ||||| ||||| |||||
4205 GCAGAGCTCAAGACCTGCACACCCGCTGACGCTGCGCCAGTTGTGTTCCA 4254
106 sgluasnglyluleuglylysthrclucystriptrylleleaspcys. 122
||||| ||||| ||||| ||||| ||||| ||||| |||||
4255 CTGCAGCCGCGGTGTGGGCGGTACTGGCTGCTTCATTGTAATTGATGCCA 4304
123 .....Glnlyaspaiaululellytvglyh1s 132
||||| ||||| ||||| ||||| ||||| ||||| |||||
4305 TGTTCGAGCGCATAGACACAGACAGACGCTGAT..GTGACCGAGAC 4351
133 Asnalathrthlysglugluleuthrthmetilegluarglyglut 149
||||| ||||| ||||| ||||| ||||| ||||| |||||
4352 GTGACATCATCATCGGTGTACAGCCCACTACATGTCACAGACAGAGAT.. 4399
149 paspdluleuthrgrtyvallysalysvalysproglyaspphepetyrv 166
||||| ||||| ||||| ||||| ||||| ||||| |||||
4400 .....CAGTATACCTTCA 4412
166 alProserglythValhisAlaileglysllylleuvalleu 182
||||| ||||| ||||| ||||| ||||| ||||| |||||
4413 TCCACGAGGCACTCTGTGAGCGTGTGGGCGTGTGC..... 4447
183 ThrGlnGlnanseraprrhrrthrytyr...leutyrAsptryAsp 198
||||| ||||| ||||| ||||| ||||| ||||| |||||
4448 .....AATACCGAGGTCCCGCGCGCGCAGCCTTACACCTAATATCA 4488
198 gLyAspAlaGlu.glyLyLeuArgGluLeuHisLeuLysserile 214
||||| ||||| ||||| ||||| ||||| ||||| |||||
4489 GAACGTGCGCCAGGTGAGGCTGCGCAGCATGTCAAGAAATGAGAGCTTG 4538
215 GluValIleGluValIProserIleProGluArgHisThValHisH1SGI 231
||||| ||||| ||||| ||||| ||||| ||||| |||||
4539 AGTTCAGAGGCTTGCCAGCT.....CCAAAGCACACACTTCGAGATTC 4582
231 uGlnIleGluAspLeuethrthrrleu 241
||||| ||||| ||||| ||||| ||||| ||||| |||||

```



```

8346 .....CTGTTCCACCAAGCTGAGACAGAGCTCTTCAAGA 8307
281 lylglutylvaltyrprophelyslyslasphismelleuleupolyr 297
8306 GAGAAAGTCGAAGCTAGTGGACAGAGACGCGCTGAGACCCGTT 8257
298 glyleuglygluphelysleu...gluglytyrvalaglu.....cy 310
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
8256 GGCTCGCGCAAGTATCGCTGATGAGGAAATTGCGAGCTGAGACTACGC 8207
310 sllevalserhis 314
|||||:|||||:|||||:|||||:|||||:|||||:|||||:
8206 TCTGGCTCCAC 8194

seq.name: /cgn2.6/prodata/1/lna/5B.COMB.seq:US-08-222-617A-24

seq.documentation block:
: Sequence 24, Application US/08222617A
: Patent No. 5882879
: GENERAL INFORMATION:
: APPLICANT: Veenstra, Annemarie E.
: APPLICANT: Martin, Juan F.
: APPLICANT: Garcia, Bruno D.
: APPLICANT: Gutierrez, Santiago
: APPLICANT: Barredo, Jose L.
: APPLICANT: Von Doehren, Hans
: APPLICANT: Palisa, Harriet
: APPLICANT: Van Lempt, Henk
: TITLE OF INVENTION: A Method for Influencing Beta-Lactam
: TITLE OF INVENTION: Antibiotic Production and for Isolation of Large
: NUMBER OF SEQUENCES: 27
: CORRESPONDENCE ADDRESS:
: ADDRESSEE: McDonnell Boehnen Hulbert & Berghoff
: STREET: 300 South Wacker Drive
: CITY: Chicago
: STATE: Illinois
: COUNTRY: USA
: ZIP: 60606
: COMPUTER READABLE FORM:
: MEDIUM TYPE: Floppy disk
: COMPUTER: IBM PC compatible
: OPERATING SYSTEM: PC-DOS/MS-DOS
: SOFTWARE: Patent Release #1.0, Version #1.25
: CURRENT APPLICATION DATA:
: APPLICATION NUMBER: US/08/222.617A
: FILING DATE: 04-APR-1994
: CLASSIFICATION: 435
: ATTORNEY/AGENT INFORMATION:
: REFERENCE/DOCKET NUMBER: 97,157
: INFORMATION FOR SEQ ID NO: 24:
: SEQUENCE CHARACTERISTICS:
: LENGTH: 11601 base pairs
: TYPE: nucleic acid
: STRANDEDNESS: single
: TOPOLOGY: linear
: MOLECULE TYPE: DNA (genomic)
: ORIGINAL SOURCE:
: ORGANISM: Acremonium chrysogenum
: FEATURE:
: NAME/KEY: CDS
: LOCATION: 388..11526
: OTHER INFORMATION: /function= "Enzyme"
: OTHER INFORMATION: /product= "Acv Synthetase"
: NAME/KEY: CDS
: LOCATION: 8050..8052
: OTHER INFORMATION:
: OTHER INFORMATION: /note= "NNN-AGU, AGC, UCU, UCC, UCA, or UCG"
US-08-222-617A-24

```

```

alignment_scores:
    Quality: 84.00      Length: 355
    Ratio: 0.503      Gaps: 25
    Percent Similarity: 47.042      Percent Identity: 23.380

alignment_block:
US-09-462-846-2 x US-08-222-617A-24/rev ..
Align seg 1/1 to reverse of: US-08-222-617A-24 from: 1 to: 11601

34 ThrGlyGluCysTTPAlaPheAlaAlaHisGlnAsnGlyGlnSerValVa 50
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
9133 ACGGGGAGAGGTGGCGTGCAGGAGGTGCAC.....GGATTCGATGCT 9090
50 lGlnAsnGlyMetTyrLysGlyPheThrLeuSerGluLeuTyr..... 64
|||||:|||||:|||||:|||||:|||||:|||||:|||||:
9089 TCGCACACAGATTCCGGCGTTGAGCGTGTAGTACGACGATGGCGCTGAC 9040
65 .....GluHisHisArgHisLeuPheGly..... 72
|||||:|||||:|||||:|||||:|||||:|||||:|||||:
9039 CCTCGTGAATCGAGCCATCTCTCATCTTGGGATGACGCGACGCT 8990
73 .....GlnLeuGluGly...AspArgPheProLeuThrLysIleL 86
|||||:|||||:|||||:|||||:|||||:|||||:|||||:
8989 CCTTAACAGCAGAGATGGCTAGACGGGCAATCCGCACTCGAGAG..C 8943
86 euAspAlaAspGln.....AspLeu..... 92
|||||:|||||:|||||:|||||:|||||:|||||:|||||:
8942 TCGATGCGGAACCCCTCAGCTTGCCTGTCTGCTTCCGGCGGAGATA 8893
93 .....SerValGlnValHisProAsnAspGluTyrAlaAs 104
|||||:|||||:|||||:|||||:|||||:|||||:|||||:
8892 CTCGAGATGGTGGGTCCACGAGAGCCAC.....CAGATTCACAG 8852
104 nileHisGlnAsnGlyGluLeuGlyLysThr...GluCysTTPtyrIleI 120
|||||:|||||:|||||:|||||:|||||:|||||:|||||:
8851 TCTGTAGAGTCT.....CTGGGACAGACTGTCACCTTGGCTGTGGCG 8808
120 leAspCysGlnLysAspAlaGluIleLeuTyrGlyHisAsn..... 133
|||||:|||||:|||||:|||||:|||||:|||||:|||||:
8807 TCGTAGAAAGGTTGGGATGATGATCGTCTGTCAGGGCATCTGTGT 8758
134 .....AlaThrThrLysGluGluLeuThrThrIle 143
|||||:|||||:|||||:|||||:|||||:|||||:|||||:
8757 GAGTACCCGCGAGACGACGATCAGCCGCCAAGTAGAGTCCGACACAG 8708
143 ttleGluArgGlyGluTyrPaspGluLeuLeuArgArgValLysValLysP 160
|||||:|||||:|||||:|||||:|||||:|||||:|||||:
8707 CGTTGAAGAAGAACAGCTGAGTGGTCTGTTCAGGAGCTGAGCTACTT 8658
160 roGlyAspPhePheThrValProSerGlyThrValHisAlaIleGlyLys 176
|||||:|||||:|||||:|||||:|||||:|||||:|||||:
8657 CCGGGGAT.....CCCGTGCAGAGTCCCTGGTAA 8626
177 GlyIleLeuAlaLeuGluThrGlnAsnSerAspThrThrTyr...Ar 192
|||||:|||||:|||||:|||||:|||||:|||||:|||||:
8625 GGGGCAATCGCCCTTGAA.....CGTGTGATGATGTTGACACGCTGC 8582
192 gleuTyrAspTyrAspArgLysAspAlaGluGlyLysLeu.....A 206
|||||:|||||:|||||:|||||:|||||:|||||:|||||:
8581 TCTC.....AGTGATACCATAGCGCTTGTGATGTGCCCC 8547
206 rglGluLeu.....HisLeuLysLys.....SerIleGluVal..... 216
|||||:|||||:|||||:|||||:|||||:|||||:|||||:
8546 GCGAAGTGGAGCGCACTTCTCAAACTGATGACGCGTGAATCTCGCGC 8497
217 .....IleGluValProSerIleProGluArgHisThrVa 228
|||||:|||||:|||||:|||||:|||||:|||||:|||||:
8496 CCGACCGGTGACCATGTGAGATGCGGACAGACGAGAGCTCAATCTGCT 8447
228 lHisHis.....G 231
|||||:|||||:|||||:|||||:|||||:|||||:|||||:
8446 GCAGACCCGAGGCGTCCGCTGATGATAGATAGTTCCTCCCTGCGCGCG 8397

```



```

; TYPE: nucleic acid
; STRANDEDNESS: double
; TOPOLOGY: linear
; MOLECULE TYPE: genomic DNA
US-08-902-632-3

alignment_scores:
      Quality: 83.00      Length: 355
      Ratio: 0.532      Gaps: 21
Percent Similarity: 43.944      Percent Identity: 21.127

alignment_block:
US-09-462-846-2 x US-08-902-632-3 ..

Align seg 1/1 to: US-08-902-632-3 from: 1 to: 2325

16 ILeTrrpglglYtHrAlaLeuAlaAspRheglYtYrThrIleProSer.. 31
   ::::: ::::: ::::: ::::: :::::
1186 TTGTCGGAGAACATAGCTACCTAGATTTTAGATCCCTGTAACCTCAT 1235
   32 .....GlnArgThrIleGluC 37
1236 CATCATCACCCACAACTCTGCGCGATACGCTCAACAGAGAGA...T 1282
   37 YSTRPAlaPhe..AlaAlaHisGlnAsnGlglnSerValValGlnAsn 52
   || ::::: ||| ||::: :::::
1283 GCAGAGATATAGACCTTGCCCCACAGCTGGCCACCGCTTCGCAAGAC 1332
   53 GlyMetYrIysGlyPhe.....ThrLeuSerGluLeuTrpGlnI 66
   ::::: ||::: ||::: :::::
1333 .....TTCCGAGATATTATCCGACCTCTGAGACCTCTGAGAGA 1376
   66 SHIsArGHisLeuRheglYglnLeuGluGlyAspArgPheProLeuLeu 83
   ::::: ::::: ::::: :::::
1377 GAGCGAGAGAGTAAAGAAAGATGAAGGCCAGATTGACCCGATCGAGA 1426
   83 hTrIysIleLeuAspAlaAspGlnAspLeuSerValGlnValHisProAsn 99
   ||::: ||::: ||| :::::
1427 GGAACTCTCTGATTACAGCAG...AGGCCATCAAGATCTCGCAAC 1473
   100 AspGluTrYAlaAsnIleHisGluAsnGlglnLeuGlyIleGlyThrGlyC 116
   ::::: ||| ||| :::::
1474 AGCTACTTC.....GGTACTACGGCTATGCAAGGCG 1505
   116 s...TrpTrYIleIleAspCysGlnIysAspAlaGluIleIleYrGln 132
   ::::: ::::: ||::: :::::
1506 GCGCTGATC.....TGCAGGAGTGTCGAGAGACGTA..... 1539
   132 ISAsnAlaThrThrIysGluGluLeuThrThrMetIleGluTrpGlyGlu 148
   ::::: ::::: ||::: ||:::
1540 ..ACGGCTGGGGAAGGAGTACATAACGATGACATCAAG..... 1578
   149 TrpAspGluLeuLeuAlaArgValIysValys..... 159
   ||::: ||::: |||
1579 ....GAGATAGAGAAAGTACGCGCTTAAAGTAACTACAGCGACAC 1622
   160 ProGlyAspRheTrYValProSerGlyTrhValHisAlaIleGlyL 176
   ||| ||| ::::: :::::
1623 GCAGCGATTTTGGCAACATACCTGGAGCGGATGCTGAACCGTCANA 1672
   176 YS.....GlyIle 178
   .....
1673 AGAAGGCTATGAGCTTCCTCAACTATATCAAGCCAAACTCCGGGCGG 1722
   179 LeuAlaLeuGluTrGln..... 184
   ||| ||||| ::::
1723 CTTGAGCTCGAGTACGAGGCGTCTTACAACGCGGCTTCTGTCAGAA 1772
   185 .....GlnAsnSerAspTrhTrYrArgLeuT 194
   ::::: ::::: ||| |||
1773 GAAGAGATGTCGGTGATAGACGAGGAAGCAAGATTAACAACGCCGCGAC 1822

```

```

194 yAspTyAspArgLysAsp..... 200
    :||:|||||
1823 TTGAGATTGTGAGCGGTGAGCGAGATACGAAAGACGAGCGG 1872
201 .....AlaGluGlyLysLeuArgGluLeuHisLeuLysSerIleG1 215
    :||:|||||
1873 AGGTTCTTAAGCTTGTCTAAAGAGCGGTGACGTGAAGAGCGGTGAG 1922
215 uValIle.....GluValProSerI 222
    :||:|||||
1923 GATAGTCAAGAGTACCGAAAGCTGACAGACAGTACGAGGTTCGG.... 1968
222 leProGluArgHisThrValHisIleGluGlnIleGluAspLeuThr 238
    :||:|||||
1969 ..CGGGAAG..CTGCTATCCACGACGATTAACGAGGATTTAAAG 2013
239 ThrThrIleuIleGluCysAlaTyrPheSerValGlyLysTrpAsnLeu 255
    :||:|||||
2014 GACTACAGCGACCGGTCCCGACGTTGCGGTGCCAAGAGGTGGCGC 2063
255 rGlySerLaserLeuLysGlnGlnLysProPheLeuLeuIleSer... 270
    :||:|||||
2064 GAGAGGAGTCAAAATTA.....CGCCCTGGAACGGGTGATTAACCTACA 2104
271 ..ValIleGluGlyLysArgMetIleSerGlyLysValTyrPro 286
    :||:|||||
2105 TCGTCTCAAGGCTGTGGAGATA.....GGCAGACAGGCGATACCG 2148
287 PheLysLysGlyAsp 291
    :||:|||||
2149 TTCAGCGAGTTCGAC 2163

seq_name: /cgn2.6/prodata/1/ina/6A_COMB.seq.us-08-834-306-17
seq_documentation_block:
; Sequence 17, Application US/08834306
; Patent No. 6054135
; GENERAL INFORMATION:
; APPLICANT: Reed, Steven G.
; APPLICANT: Skeiky, Yasir A.W.
; APPLICANT: Lodes, Michael J.
; APPLICANT: Houghton, Raymond L.
; TITLE OF INVENTION: COMPOUNDS AND METHODS FOR THE DETECTION AND
; NUMBER OF SEQUENCES: 65
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: SEED and BERRY LLP
; STREET: 6300 Columbia Center, 701 Fifth Avenue
; CITY: Seattle
; STATE: Washington
; COUNTRY: USA
; ZIP: 98104-7092
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/834,306
; FILING DATE: 15-APR-1997
; CLASSIFICATION: 424
; ATTORNEY/AGENT INFORMATION:
; NAME: Maki, David J.
; REGISTRATION NUMBER: 31,392
; REFERENCE/DOCKET NUMBER: 210121.422C1
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (206) 622-4900
; TELEFAX: (206) 682-6031
; INFORMATION FOR SEQ. ID NO.: 17:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 2446 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear

```

us-08-834-306-17

```

alignment_scores:
  Quality: 82.50      Length: 262
  Ratio: 0.693       Gaps: 15
  Percent Similarity: 45.420  Percent Identity: 20.992

```

```

alignment_block:
US-09-462-846-2 x US-08-834-306-17 ..

```

```

Align seg 1/1 to: US-08-834-306-17 from: 1 to: 2446

```

```

26 GlyTyrThrIleProSerGlnArgThrGlyLysTrpAlaPheAla1 42
    :||:|||||
1359 GGAGTGAAGTTTGAAGCAGCCGAGCAGGCGGCGGCGCTTGGCC.. 1406
42 HisGlnAsnGlyIleSerValValGlnAsnGlyMetTyrLysGlyPhe 59
    :||:|||||
1407 ....CGAAGTGACAGATCAACCGTACATTTCTTACACAAAGCTTCA 1452
59 hrLeuSerGluLeuTrpGluHisArgHisLeuPheGlyLeuGlu 75
    :||:|||||
1453 CTGTAGTGTGATGCGGTGATCCAC..... 1478
76 GlyAspArg.....PheProLeuLeuThrLysIleLe 86
    :||:|||||
1479 ..GATAGCGCGAAGAAACGACCCCATTCCTTTGATTCGTGTGAT 1525
86 uAspAlaAspGlnAspLeuSerVal.....GlnValHisP 98
    :||:|||||
1526 GGAT...GACATGACAAAGACTGTGCTATTGCTGTTTACACCCATG 1572
98 roAsnAspGluTyrAlaAsnIleHisGlnAsnGlyLysGlyLysThr 114
    :||:|||||
1573 AAGGAGAGTGGATGATCTGATATTCATGTCGCGTACAAATACTTTCA 1622
115 GluCysTrpTyrIleIleAspCysGlnLysAspAlaGluIleTyrG1 131
    :||:|||||
1623 ACAGGCTGTG.....GACCCGAGAAACCGCTGCAGTGTGCTGCG 1663
131 yHisAsnIleThrThrLysGluGluLeuThrThrMetIleGluArgGly 148
    :||:|||||
1664 ACAC.....GACACGGGCGC 1677
148 IuTrpAspGluLeuLeuArgArgValLysValLysProGlyAspPhe 164
    :||:|||||
1678 ATTGGGAT.....TTC 1688
165 TyrValProSer.....GlyThrValHisAlaIleGly 176
    :||:|||||
1689 TACGTTAAGCGGAGGAGCTTACTTGGACCTAC.....AA 1726
176 sGlyIleLeuAlaLeuGluThrGln...GlnAsnSerAspThrThrTyrA 192
    :||:|||||
1727 GGGTCTCTTCCAAACAAACAGATATTACATCATCAATTCACGGGGA 1776
192 rGluTyrAspTyrAspArgLysAspAlaGluGlyLysLeuArgGlu... 207
    :||:|||||
1777 GAGTG.....GGGAGGTTCAGAGTCCA 1799
208 .....LeuHisLeuLysSerIleGluValIleGlu.....ValPr 220
    :||:|||||
1800 GCCATTTCACCTCTTCAAGCCCGTTGTGATTAACGAGACTCAATTC 1849
220 oSerIleProGluArgHisThrValHisIleGluGlnIleGluAspLeu 237
    :||:|||||
1850 AAGCTTAAGATGCTGATGCTGGGAGACAGCGCCAAATTAACATGA 1899
237 eutThrThrLeuIle...GluCysAlaTyrPheSerValGlyLysTrpAs 253
    :||:|||||
1900 TGGACGCTTATTCTCGAGTGTGCGTCCGAGAAAGACAGAGTGGCA 1949

```



```

3213 GGTGCTGATACATCCTTCGTGTGACAGCGTAC..... 3245
303 sleuGluclyThrAlaGlu...CysIleValSerHis 314
|||||.....|
3246 .AAGAGGCGCTACGAGAGAGATGCGTAACGATCCAT 3281
seq_name: /cgn2_6/ptodata/1/lna/5A_COMB.seq:US-08-961-083-105

seq_documentation_block:
; Sequence 105, Application US/08961083
; Patent No. 6159469
; GENERAL INFORMATION:
; APPLICANT: Chol et. al.
; TITLE OF INVENTION: Streptococcus pneumoniae Antigenes and Vaccines
; NUMBER OF SEQUENCES: 452
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Human Genome Sciences, Inc.
; STREET: 9410 Key West Avenue
; CITY: Rockville
; STATE: Maryland
; COUNTRY: USA
; ZIP: 20850
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette, 3.50 inch, 1.4mb storage
; COMPUTER: HP Vectra 486/33
; OPERATING SYSTEM: MSDOS version 6.2
; SOFTWARE: ASCII Text
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/961,083
; FILING DATE:
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER:
; FILING DATE:
; ATTORNEY/AGENT INFORMATION:
; NAME: Brookes, A. Anders
; REGISTRATION NUMBER: 56,373
; REFERENCE/DOCKET NUMBER: PB340P2
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (301) 309-8504
; TELEFAX: (301) 309-8512
; INFORMATION FOR SEQ ID NO: 105:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1879 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: double
; TOPOLOGY: linear
; US-08-961-083-105

alignment_scores:
Quality: 81.00 Length: 341
Ratio: 0.570 Gaps: 11
Percent Similarity: 41.642 Percent Identity: 17.889

alignment_block:
US-09-462-846-2 x US-08-961-083-105 ..
Align seg 1/1 to: US-08-961-083-105 from: 1 to: 1879

26 glyTyrThrIleProSerGlnArgThrGlyGluCysTrpAlaIleAla 42
|||.....|
23 GGTCAAGCTTACCTGAAGACATCGGAAACGAAAGCGTACTTATC 72
|||.....|
42 aHisGlnsnGlyInsValValGlnAsnGlyMetTyrGlyGlyPhe 59
|||.....|
73 AGAAAGACGAGAGACCGTTCTACTCAAGCAAACTGAGGCGTTA 122
|||.....|
59 hr..... 59
|||.....|
123 CTGAATAACGATTCACCTCCGACACCTACAGAAAGAACTGAAGTAG 172
|||.....|
60 .....LeuSerGluLeuTrpGluHisHisArgHis 69

```

```

173 GAGGAACAAAGCCCTTCTAGCTGATACACTTTTGAAGAAAGATGA... 220
|||...|
69 sleuPheGlyGlnleuGluValAspArgPheProLeuLeuThrGlyIle 86
|||...|
221 .....CAGCTCAAAAATAATCCAGAGCTTACAGATGTCT 254
|||...|
86 eu.....AspAlaAspGlnAspLeuSer 93
|||...|
255 TAAAGAAACTGTAGATACAGCTGATGTGATGGACACAGCAAGTCCA 304
|||...|
94 ValGlnValHisProAsnAspGluTyrAlaAsnIleHisGlnAsnGly 110
|||...|
305 GAGAAATACTCTCTGACACAGTAAAGGTGAGTGAAGAAATATACAA 354
|||...|
110 uleuGlyLysThrGluCysTrpTyrIleLeaAspCysGlnLysAspAla 127
|||...|
355 A.....GACAGCATGATGTTC 371
|||...|
127 IuIleIleTyrGlyHisAsnAlaThrThrLysGluGluLeuThrThrMet 143
|||...|
372 CTGCTGCTTATCTTGAAGAAAGCTGAAGGAAAGCTCTTCACTGCCGT 421
|||...|
144 Ile.....GluArgGlyGluTrpAspGluLeuLeu 154
|||...|
422 GTAAACCAAGTAATCTCTTATGAACTATTCGCGTGAATGATGTATAC 471
|||...|
154 garValLysValLysProGlyAsp..... 162
|||...|
472 TCGCTATTACTTAAAGCTTCGGTAAATGCTCCTGTGTCGACAAATGTA 521
|||...|
162 ..... 162

522 CTGCTAAATAATCCTTCCTTACCTCCTTGAAGATTAAACAAAGGANA 571
|||...|
163 PhePheTyrValProSerGlyThrValHisAlaIleGlyLysGlyIle 179
|||...|
572 TACTTCTATGAGTAACTTAAATGCGCAATCTGTGTAACAAAGGTCA 621
|||...|
179 uAlaLeu.....GluThrGlnGlnAsnSerAspThrThrTyrArg.... 192
|||...|
622 AGCTTATATGATCAACTTCGCGGTATGTACTCAACTATTAAGCTA 671
|||...|
193 ..LeuTyrAspTyrAspArgLysAspAlaGluGlyLysLeuArgGluLeu 208
|||...|
672 CTGTTAAGTTTACGAAATTAAGACGCTGAAGCTGACTGACTAATCTA 721
|||...|
209 HisLeuLysLysSerIleGluValIleGluValProSerIleProGlu 225
|||...|
722 GTTGTACTTAATAATGTAGAC...ATCAACATCATGATGTACTTCTTA 768
|||...|
225 gHisThrValHisHisGluGlnIleGluAspLeuLeuThrThrLeu 242
|||...|
769 AGAAGACGTTCAAAAAGCCGTTGCAGACACCTTAAAGACAGTATGAG 818
|||...|
242 legLucysAlaTyrPheSerValGlyLysTrpAsnLeuSerGlySerAla 258
|||...|
819 TTCAGACAGCTTACTAGAAAAAGCCAG..... 847
|||...|
259 SerLeuLysGlnLysProPheLeuLeuIleSerValIleGluGly 275
|||...|
848 .....GTTGA 852
|||...|
275 uGlyArgMetIleSerGlyGluTyrValTyrProPheLysLysGlyAsp 292
|||...|
853 AGGT.....CCATTACAGCAGAGTGTCA 875
|||...|
292 IsMetLeuLeuProTyrGlyLeu 299
|||...|
876 ACCATGTGATTCATACGAACTC 898
|||...|
seq_name: /cgn2_6/ptodata/1/lna/5A_COMB.seq:US-08-040-751-4

```

```

seq documentation block:
; Sequence 4, Application US/08040751
; Patent No. 5407825
; GENERAL INFORMATION:
; APPLICANT: Payne, Jewel M
; APPLICANT: Sick, August J
; TITLE OF INVENTION: No. 5407825el Bacillus thuringiensis isolates
; TITLE OF INVENTION: active against Lepidopteran Pests and Genes Encoding No. 5407825el
; TITLE OF INVENTION: Lepidopteran-active Toxins
; NUMBER OF SEQUENCES: 4
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: DAVID R. SALIWANCHIK
; STREET: 2421 N.W. 41st STREET, SUITE A-1
; CITY: GAINESVILLE
; STATE: FL
; COUNTRY: USA
; ZIP: 32606
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patentin Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/040,751
; FILING DATE: 19930329
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: SALIWANCHIK, DAVID R.
; REFERENCE/DOCKET NUMBER: MA39.C1.D3
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 904-375-8100
; TELEFAX: 904-372-5800
; TELEX:
; INFORMATION FOR SEQ ID NO: 4:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 3522 base pairs
; TYPE: NUCLEIC ACID
; STRANDEDNESS: double
; TOPOLOGY: linear
; MOLECULE TYPE: DNA (genomic)
; HYPOTHEICAL: NO
; ANTI-SENSE: NO
; ORIGINAL SOURCE:
; ORGANISM: Bacillus thuringiensis
; STRAIN: aizawai
; INDIVIDUAL ISOLATE: PS81a2
; IMMEDIATE SOURCE:
; LIBRARY: Lambdaagem - 11 (tm) Library of August Sick
; CLONE: 81A2
; US-08-040-751-4

alignment_scores:
Quality: 81.00 Length: 347
Ratio: 0.529 Gaps: 19
Percent Similarity: 44.092 Percent Identity: 19.885

alignment_block:
US-09-462-846-2 x US-08-040-751-4 ..
Align seg 1/1 to: US-08-040-751-4 from: 1 to: 3522

54 MetYrYlVSGlypHeThrLeuSerGIuLeuTrpGIuHISIArGHISle 70
+ ::::::::::: ||:::::::::
2173 TTGATTCAAAAAATAGATGAGTCGAATTTAAAGCCTATAACCGTTAC.. 2220
+ ::::::::::: ||:::::::::

70 uPheGlyInLeuGIuGIuYAspArgPheProLeuLeuThrYlSleLeuA 87
+ ::::::::::: ||:::::::::
2221 .....CAATTAAGAGG.....TATATCG 2239
+ ::::::::::: ||:::::::::

87 sPAlAspGInAspLeuSer.....ValGInValHISProAsnAsp 100
+ ::::::::::: ||:::::::::
2240 AAGATAGTCATAGACTAGAAATCTATTATTATTCGCTACATGCATAAACAC 2289
+ ::::::::::: ||:::::::::

```

101 GtuyAlaAsnIleHisGlnGlyLeuGlyLysThrGluCysPsr 117
||| :|||::|||::|||
2290 GAACAGTAATGTATCCAGCGTACGGGTCTTA.....TG 2324

117 PylrIle..... 119
| : :
2235 GCCGGTTTCAGTCGAAGTCCAAATTGGAAGGTGTGGAGAACCGAATCGGT 2374

120 IleasprCysGln... Lysasp 125
2375 GTGTGCACACCTTGAAATGGAATCCGTGATTTAAGTGTCTTCGCAAGAC 2424

126 AlaGlutIleIleTyrglyHisAsn..... 133
:||| | |||:::
2425 GGCGAAAAATGTGCCAATCATCTCCATCATTTCTCCCTGGACAFTGATG 2474

134 ... AlaThrThrylsGlnGlyLeuThrThrMetIleGluArgGlyLut 149
||| :|||
2475 TGGATGTGCAACTGTCCGAAGG..... GATCTAGCGGTG 2509

149 rPaspglIleuenuArgrVallysValysProGlyAspPhePheTyR 165
|| : : : : :
2510 GG..... GTGTATATCAAGATTAAAGCCAGGAAGT..... Tat 2544

166 ValProSerGlyThrValHis..... AlaIleGlyLy 176
||| :|||:::
2545 GCAGAGTTTGGAAATCTGGAATTTATCGAAGAGAACCATTAAATGGAGA 2594

176 sglYile..... 178
: : : : :
2595 AGCAGCTGCTGTGGAAAGAGCGGAAAAAATGAGACACAACCGG 2644

179 LeuAlaLeugluThr..... GlnGlnAsn 186
||| |||||
2645 AAAAATACAAATTGAAAACAAAAGATATATACAGAGCAAAGAGCT 2694

187 SerAspThrThrTyArgLeutyrrAspIyrrAsparGlyAsaParlaGluGl 203
||| :|||:::
2695 GTGGATGCTTATTCGTAGATTCCTCAATATGTAAATGATTACAGACGATAC 2744

203 LysIleu..... ArgGluIleuHisIleuLysLys 213
: : : : :
2745 AAACATTTGGATGATTCATCGCGGACAGATAGACTTTCATCAGATTCACG 2794

213 erIleGluValIleGluValProSerIleProGluArgHisThrValHis 229
: : :|||::|||
2795 AGGCTATTCCTCCAGAACATCTTCATTCACGGAATAAATGTGGTATT 2844

230 HisGlnGlnIleGluAspIleuThrThrThrIleu..... 241
||| :|||::|||
2845 TTTCGAATAATTACAAAACCGTATTTTTCATCGATTATCCCTATATGATGC 2894

242 IleGluCysAlaTyrrPheSerValGlyLys..... Trpa 253
||| :|||:::
2895 GAGAAATGTCATTAATAAATGGCATTTCAATATGAGCTTATCATGCTGGA 2944

253 snIeuSerGlySerAlaSerLeuLysGlnGlnLysProPheIleuIle 269
: : :|||::|||
2945 ACCGTGAAGAGCATGTGTGATGTGTGAACMAAACMAACCAACCGTTGGGC 2994

270 SerValIle.. GluGlyGlnGlyArgMetIleSerGlyGluTyrrValTy 285
||| :|||
2995 CTGTGTGTCGCCGAATGGAGACAGAAAGTGTCCAACAACATTCGTGTGCG 3044

285 rProPheLysLysGlyAspHisMetIleuProTyrglyLeuGlyLup 302
||| :|||
3045 TC CGGGGGCGGTATATCTCCGTGTTCACAGCGTAC..... 3081

302 heLysLeuGlnGlyTyrrAlaGlu... CystLevaSerHis 314
||| :|||
3082 AAGAGGGATATGAGGAAGGTTGGCTATACATTCAT 3117


```

||||: ||| |||:||||: |||
2995 CTTGTTGTCCTCCGATGCGAAGCAGAACTGTCACAAACAAATCGTGTCTG 3044
285 rProphelyslysglyAspHisMetLeuProTyrGlyLeuGlyGluP 302
||| :|||
3045 TCCGGGGCGGTGGCTATATCTCCGCTTACAGCGTAC..... 3081
302 helysLeuGluGlyTyrAlaGlu...CysIleValSerHis 314
|||||:||||| |||:|||||
3082 .....AAAGAGGAGATGAGAGAGTTGCGTAAACCATCCAT 3117

```


XX Disclosure; Fig 1A-B: 31bp; English.

XX This sequence represents the Bacillus subtilis cysteine protease CP1
 CC coding sequence. The invention relates to new Gram-positive bacteria,
 CC especially Bacillus, containing mutations in, or deletion of, at least
 CC part of the gene encoding cysteine protease (CP1, CP2 (X25734) or CP3
 CC (X25735), so that CP-related proteolytic activity is lost. The bacteria
 CC are used for expressing heterologous proteins, particularly hormones,
 CC growth factors, cytokines or especially enzymes, e.g. protease,
 CC carboxypeptidase, lipase, isomerases (racemase, epimerase, tautomerase or
 CC mutase), transferase, kinase and phosphatase. CP1-3 are used e.g. in
 CC soaps, dishwashing compositions, contact lens cleaners or laundry
 CC detergents, also for peptide hydrolysis, waste or textile treatment,
 CC to cleave fusion proteins and as animal feed additive.

XX Sequence 945 BP: 288 A; 180 C; 232 G; 245 T; 0 other:

Query Match 100.0%; Score 945; DB 20; Length 945;
 Best Local Similarity 100.0%; Pred. No. 1e-296;
 Matches 945; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```

Oy 1 atgacgactgaacgcttatttttcaagcctgtttcaagaagaagaatttggcgagacc 60
    |||||||
Db 1 atgacgactgaacgcttatttttcaagcctgtttcaagaagaagaatttggcgagacc 60
Oy 61 gcttaagctgatttggctataccatccgttcaacagcaggggagtgctggcttt 120
    |||||||
Db 61 gcttaagctgatttggctataccatccgttcaacagcaggggagtgctggcttt 120
Oy 121 gccgcgacataaattggtcaaaagcgtgtgttcaaaaacgggaattgaagggttcacgctc 180
    |||||||
Db 121 gccgcgacataaattggtcaaaagcgtgtgttcaaaaacgggaattgaagggttcacgctc 180
Oy 181 agcgaattatgggaacatcacacattatcgcgacgcttgaaggagaccgtttccct 240
    |||||||
Db 181 agcgaattatgggaacatcacacattatcgcgacgcttgaaggagaccgtttccct 240
Oy 241 ctgcttacaanaaatatagatgtctgcacaggaactatctgttcaggttcacgaatgat 300
    |||||||
Db 241 ctgcttacaanaaatatagatgtctgcacaggaactatctgttcaggttcacgaatgat 300
Oy 301 gaatatcccaacatcacatgaacacgcttgagcttgaaaaaacgaatgctgttaccattat 360
    |||||||
Db 301 gaatatcccaacatcacatgaacacgcttgagcttgaaaaaacgaatgctgttaccattat 360
Oy 361 gattgcacaaaagaatgcgagatatttattgscacaatgacacacaaaggaagaacta 420
    |||||||
Db 361 gattgcacaaaagaatgcgagatatttattgscacaatgacacacaaaggaagaacta 420
Oy 421 actaccatgataagcgttggaagatggaatgagctcttcgcgcgtgtataaagtaacgcg 480
    |||||||
Db 421 actaccatgataagcgttggaagatggaatgagctcttcgcgcgtgtataaagtaacgcg 480
Oy 481 gggggaattttctatgtgcacacgggttactgttcattgagatgggaaaggaaattcttgc 540
    |||||||
Db 481 gggggaattttctatgtgcacacgggttactgttcattgagatgggaaaggaaattcttgc 540
Oy 541 ttggaacgcgacgagaactcagacacaaactacagattatgatataagacggaagaat 600
    |||||||
Db 541 ttggaacgcgacgagaactcagacacaaactacagattatgatataagacggaagaat 600
Oy 601 gcaagaagcgaacgctgcgagcttcctcgaaaaagaagattgaagtgatagagttcccg 660
    |||||||
Db 601 gcaagaagcgaacgctgcgagcttcctcgaaaaagaagattgaagtgatagagttcccg 660
Oy 661 tctattccagaagcgcatacagttcacatgaacaaatgagatttcttacaacgaca 720
    |||||||
Db 661 tctattccagaagcgcatacagttcacatgaacaaatgagatttcttacaacgaca 720
Oy 721 ttgattgaatgcgttacttttcggtgggaaatggaacttatcaggtcgcgaagctta 780
    |||||||

```

```

Db 721 ttgattgaatgcgttactttcgttggggaatggaacttatcaggtcgcgaagctta 780
Oy 781 aagcagcaaaaacacatctcttctatcagttgattgaaggagcgctatgactct 840
    |||||||
Db 781 aagcagcaaaaacacatctcttctatcagttgattgaaggagcgctatgactct 840
Oy 841 ggtgaatgtctatctcttcaaaaagagatcatatgttgcgtcctcagcttga 900
    |||||||
Db 841 ggtgaatgtctatctcttcaaaaagagatcatatgttgcgtcctcagcttga 900
Oy 901 gaatttaactcgaagatgatacgaatgatatgtcttcccatctg 945
    |||||||
Db 901 gaatttaactcgaagatgatacgaatgatatgtcttcccatctg 945

```

RESULT 2
 X25734
 ID X25734 standard; DNA; 945 BP.

XX X25734:
 XX 21-MAY-1999 (first entry)

XX B. subtilis cysteine protease CP2 coding sequence.

KW Cysteine protease; Gram-positive bacteria; mutation; deletion;
 KW proteolysis; expression; heterologous protein; ss.

XX Bacillus subtilis.

XX W09904016-A2.

XX 28-JAN-1999.

XX 14-JUL-1998; 98WO-US14529.

XX 15-JUL-1997; 97EP-0305227.

XX (GEMV) GENENCOR INT BV.
 XX (GEMV) GENENCOR INT INC.

XX Estell DA;

XX WPI: 1999-132260/11.
 XX P-PSDB: W99372.

XX Gram-positive bacterium with mutated or deleted gene for cysteine
 XX protease 1, 2 or 3 - used to express proteins with reduced
 XX proteolytic degradation, e.g. proteins, growth factors or enzymes

XX Disclosure; Fig 5A-B: 31bp; English.

XX This sequence represents the Bacillus subtilis cysteine protease CP2
 CC coding sequence. The invention relates to new Gram-positive bacteria,
 CC especially Bacillus, containing mutations in, or deletion of, at least
 CC part of the gene encoding cysteine protease (CP1 (X25733), CP2 or CP3
 CC (X25735), so that CP-related proteolytic activity is lost. The bacteria
 CC are used for expressing heterologous proteins, particularly hormones,
 CC growth factors, cytokines or especially enzymes, e.g. protease,
 CC carboxypeptidase, lipase, isomerases (racemase, epimerase, tautomerase or
 CC mutase), transferase, kinase and phosphatase. CP1-3 are used e.g. in
 CC soaps, dishwashing compositions, contact lens cleaners or laundry
 CC detergents, also for peptide hydrolysis, waste or textile treatment,
 CC to cleave fusion proteins and as animal feed additive.

XX Sequence 945 BP: 290 A; 189 C; 223 G; 243 T; 0 other:

Query Match 33.8%; Score 319; DB 20; Length 945;
 Best Local Similarity 59.4%; Pred. No. 2.2e-93;
 Matches 560; Conservative 0; Mismatches 380; Indels 3; Gaps 1;

Oy 4 agcactgaacgcttatttttcaagcctgttttcaagaagaagaatttggcgagaccgct 63

```

Db      1 atgacgcataccatttttttagagcctgtctttaaagaagacattggggaggaagaa 60
Qy      64 ttagtga---tttggtatacccttcgcgcacagaagggaggtgtggttt 120
Db      61 ctctgtagcgtcttggcagcaataccctcaaaaaaaggtgagtgctggcgct 120
Qy      121 gcgcgcatacaaatgycgaagcgttgtcaaaacggaatgataaggggttcaagctc 180
Db      121 tctgacaatgccatgctgcgtcgtctgtaaaaaatggccgcgtcgcaggaagacactt 180
Qy      181 agcgaattatgggaacatccacagacattatcggacagcgttgaaggagaccttccct 240
Db      181 gatcaagataggaagatccatccagagatccgggttccggatcgaagtggtttccg 240
Qy      241 ctcgttacaanaattagatgtcgcacagacattatcgttcaggtgcatccgaatgat 300
Db      241 ctgctggtataagctgtcgtgacgcgaatatgatactctcctcgtcgaagtcacatcgtat 300
Qy      301 gaatatgccacatacagaataaagcgtgtgagcttggaaaaacagaatgctgtacattat 360
Db      301 gattatgcaaaactgcacgaataatgycgacctgtgtaaacggaatgctgtatattcatt 360
Qy      361 gattgccaataaagatgcgcgaattttatgccaacatgcaacaaagaagaacta 420
Db      361 gatgcaaaagatgacgcgaactaatttggagacatcatgcaacaaagaagaagctc 420
Qy      421 actacatgatagagcgttggagaatggaatgagctcttcgcgcgtgttaagtgagcg 480
Db      421 aaacacagatagaagcgggtgattggaacggtcgtcgtgagcgatacaaaatcaagca 480
Qy      481 ggggattttttctatgtgcgaagcgttactgttcaatcgatggaaagaattcttgt 540
Db      481 gggagatttctttaaagtgcaagcgttaccatccacgtcttataaggaacccctgtcc 540
Qy      541 ttgagagcgaagcagacatcaacacactacagattatgatgatgacccgaagaat 600
Db      541 ctggaatccagaacactctgtacacatctcgtatcagattatgacgcgtgtaatt 600
Qy      601 gcaagaagcagcgtcgcgagcttcatctgaaaaaagatggaatgtagagtcggc 660
Db      601 gaccagggccaaaaaagaactcttcatatagaaaaaagacatggaatcacaagataccg 660
Qy      661 tctatccagaacgcgatacagcttccatccatgaaacattgagattgttacaacgaca 720
Db      661 catatcgataaagtgcatacaccggaagtaaaagaattgtagcgtgagatcattgtt 720
Qy      721 ttgattgaatgcgttacttcttcgttgggaaatggaacttaacagatcagaagctta 780
Db      721 tatgtgaatcagattattctcagtgtaacaatggaagattagcgccgagctgtctt 780
Qy      781 aagcagaacaaacacactctcttcatcagtgatggaaggagggccgtatgactct 840
Db      781 cctcatatacaaacctatttgcctgggagtgcttctgagcggatcagaacaaatcaaal 840
Qy      841 ggtgagatgtctatccttcaaaaaaagagagacataatgtctgcttcagcttctgga 900
Db      841 aatggtatcagatgtaatgcatacagcgtcacaacttattcttcgctgagcatttga 900
Qy      901 gaattaaactcgaagatatgcaagaatgatacgtctccatc 943
Db      901 gaattacaatagaagaacatgtaattcatgatattcattc 943

```

RESULT 3
 X25735
 ID X25735 standard; DNA; 948 BP.
 XX
 AC X25735;
 XX
 DT 21-MAY-1999 (first entry)
 XX
 DE B.subtilis cysteine protease CP3 coding sequence.

```

XX      KW Cysteine protease; Gram-positive bacteria; mutation; deletion;
KW      KW proteolysis; expression; heterologous protein; ss.
XX      OS Bacillus subtilis.
XX      PN M090904016-A2.
XX      PD 28-JAN-1999.
XX      PF 14-JUL-1998; 98MO-U0514529.
XX      PR 15-JUL-1997; 97EP-0305227.
XX      PA (GENEV) GENENCOR INT BV.
XX      PA (GENEV) GENENCOR INT INC.
XX      PI Estell DA;
XX      DR WPI: 1999-132260/11.
XX      DR P-PSDB: W99373.
XX      PT Gram-positive bacterium with mutated or deleted gene for cysteine
PT      PT protease 1, 2 or 3 - used to express proteins with reduced
PT      PT proteolytic degradation, e.g. proteins, growth factors or enzymes
XX      PS Disclosure; Fig 6A-B; 31pp; English.
XX      CC This sequence represents the Bacillus subtilis cysteine protease CP3
CC      CC coding sequence. The invention relates to new Gram-positive bacteria,
CC      CC especially Bacillus, containing mutations in, or deletion of, at least
CC      CC part of the gene encoding cysteine protease (CP)1 (X25733), CP2 (X25734)
CC      CC or CP3, so that CP-related proteolytic activity is lost. The bacteria
CC      CC are used for expressing heterologous proteins, particularly hormones,
CC      CC growth factors, cytokines or especially enzymes, e.g. trypsinase,
CC      CC carboxypeptidase, lipase, isomerases (racemase, epimerase, tautomerase or
CC      CC mutase), transferase, kinase and phosphatase. CP1-3 are used e.g. in
CC      CC soaps, dishwashing compositions, contact lens cleaners or laundry
CC      CC detergents, also for peptide hydrolysis, waste or textile treatment,
CC      CC to cleave fusion proteins and as animal feed additive.
XX      CC
XX      CC Sequence 948 BP; 286 A; 192 C; 244 G; 226 T; 0 other;
SQ

```

Query Match 33.0%; Score 311.8; DB 20; Length 948;
 Best Local Similarity 58.8%; Pred. No. 47e-91;
 Matches 557; Conservative 0; Mismatches 387; Indels 3; Gaps 1;

```

Qy      1 atgacgactgaacgcttattttcaagcctgtttcaagaagaattggggcgagacc 60
Db      1 atgacgcaatccacgatttttctaagcctgtgtttaagaataatctggggcgagacc 60
Qy      61 gctttagctgat---ttggctataccatctccgtacacgaacgaacgggagtgctggct 117
Db      61 gctttagagatagatttggatacagatctctctcattagatcaacggggagtgctggcc 120
Qy      118 ttggcgcgcatacaaatggtcaaaagcgttgttcaaaacggaatgataaagggttcaag 177
Db      121 atttcgcctacccaagaagcagcagcgtgtgtaaaatgagccgtatataaaggaaagaca 180
Qy      178 ctaagcgaattatgggaataacatcacagacattatcgcgaacgcttgaaggggagcgttct 237
Db      181 ttgatcgagcttgggaagagacgcgtggaatcttcggcggcgttagaaggagatccgctt 240
Qy      238 cctcgtctacaanaattatagatgtgcacgaactatcgttcaggtgcatccgaat 297
Db      241 ccgctcttcaaaaagcgtgtgtgtaaggaagatacgtcaatataaagttcacctgat 300
Qy      298 gatgaatatgcgaacatacatacagaacagcgttggaagaaacagaatgctgtgtaatt 357
Db      301 gattactatgcgcgaagaacgaagaggaactcgcgaagcgaagatgctgtgtaatt 360
Qy      358 attgatgccaanaaagatgacgagattattatggtccacaatgcaacaacaaagaaga 417

```

```

Db      361 atgacgtatgagaaacgcagaaatcattctcggcagacgcccgcgtcaaaacgaa 420
Qy      418 ctaactaccatagacgttgagaaatgagatgctcttcgcccgttaagtgaa 477
Db      421 ctgtcccaatgtatcaacgcgtgtgagggcctctctcgaagaatcaaatlaa 480
Qy      478 ccgggagattttctatgtgccaacgcgttactgttcatgtcagatcgaaagaaatctt 537
Db      481 ccgggagattttctatgtgccaacgcgttactgttcatgtcagatcgaaagaaatctt 540
Qy      538 gctttgagacgcagcagacatcaactacacatcagattatgattatgaccca 597
Db      541 gttttgagacgcagcagacatcaactacacatcagattatgattatgaccca 600
Qy      598 gatgcaaaagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagc 657
Db      601 gatgcaaaagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagc 660
Qy      658 ccgtctcttcagaaacgcagcagcagcagcagcagcagcagcagcagcagcagcagc 717
Db      661 ccccatgtgacgcgtatagatagatgacagacagacagacagacagacagacagac 720
Qy      718 acattgattgaatgcgtcattcttcgtgtgagaaatgaaatgaaatgaaatgaaatgaa 777
Db      721 acattgtccaagggagaaatcttctcgtttataaattgagacatcaatgagcagcagc 780
Qy      778 ttaagacgacaaacacatcctctctatcagtgatgagtgagggagggccgtatgac 837
Db      781 atgctatgagatgacatcctctctatcagtgatgagtgagggagggccgtatgac 840
Qy      838 tctgtgagatgtctatcctctctcaaaaagagatcatatgtctgccttaagcgtct 897
Db      841 tatgagacaaacacatcctctctcaaaaagagatcatatgtctgccttaagcgtct 900
Qy      898 ggaagaatttaactcgaagatgtcagaatgtatcgtctcccatc 944
Db      901 ccgattttacgataaaagaaactgttacccttatcgttctcatat 947

RESULT 4
ID      X13110 standard; DNA; 12438 BP.
AC      X13110;
DT      19-MAR-1999 (first entry)
DE      Enterococcus faecalis genome contig SPQ ID NO:173.
KW      Enterococcus faecalis; contig; detection; Enterococcal infection;
KM      vaccine; attenuation; computer readable medium; ds.
OS      Enterococcus faecalis.
PN      WO9850555-A2.
PD      12-NOV-1998.
PF      04-MAY-1998; 98WO-US08985.
PR      14-NOV-1997; 97US-0066009.
PR      06-MAY-1997; 97US-0044031.
PR      16-MAY-1997; 97US-0046655.
PA      (HUMA-) HUMAN GENOME SCI INC.
PI      Barash SC, Dillon PJ, Kunsch CA;
PT      WPI; 1999-045171/04.
XX      New isolated Enterococcus faecalis polynucleotides and polypeptides
XX      - used to develop products for the detection of Enterococcus and for

```

```

PT      use in vaccines for prevention or attenuation of Enterococcus
PT      infection.
XX      PS      Claim 1, Page 963-969; 2084pp; English.
XX      CC      A computer readable medium has been developed which has recorded on it
XX      CC      982 nucleotide sequences isolated from the Enterococcus faecalis genome.
XX      CC      X12938 to X13919 represent these nucleotide sequences which are primarily
XX      CC      nucleotide sequences, also known as contigs. The computer-based system
XX      CC      can identify fragments of the Enterococcus faecalis genome with
XX      CC      commercial importance. The products can be used to detect the presence
XX      CC      of Enterococcus faecalis in samples. They can also be used for
XX      CC      diagnosing Enterococcal infection in an animal and monitoring
XX      CC      progression of disease, and for identifying agents which can be used to
XX      CC      modulate the growth or pathogenicity of Enterococcus faecalis, or
XX      CC      another related organism, in vivo or in vitro. In particular the
XX      CC      polypeptides encoded by the Enterococcus faecalis nucleotide sequences
XX      CC      can be used in vaccines to prevent or attenuate an Enterococcal
XX      CC      infection.
SQ      Sequence 12438 BP; 4071 A; 2218 C; 2424 G; 3708 T; 17 other:

Query Match      27.3%; Score 258.4; DB 20; Length 12438;
Best Local Similarity 58.3%; Pred. No. 4.3e-73;
Matches 492; Conservative 0; Mismatches 346; Indels 6; Gaps 2;

Qy      10 gaacgttatcttcaagcgtcttcaaaagaatcttgaggcgacgcgttg-- 67
Db      6276 gaaccatttttttaaaccttcttcaaaaagaatcttgaggcgacgtctcaac 6335
Qy      68 -cgatttggtatccatccatccgtacacacagacagggaggtgctggcctttgcccgcg 126
Db      6336 acgctatttggttattgattaccagcgataaattggtgaagattggtgcaatcagtcg 6395
Qy      127 catcaaaatggtcgaagcgtttcaaaacggaatgataaggggttcacgtcaagcgaa 186
Db      6396 catccacatggtgctgctctttaaattgagcaatttaaaagaaaaaattgatgaa 6455
Qy      187 ttatggaaacatcacagacattatctgcagacagcgttgaaggagacgttccctcgtct 246
Db      6456 ttgtggcagacatcaagaatgattatcgcgcagtgagtggtgcgtcttccattat 6515
Qy      247 acaaaaatattagatgctgacccaagactatctgtcagtggtgcacacgaatgaaat 306
Db      6516 acgaaaaatttagatgctgaagatgattatccgtgacagtcatccctgagatgacat 6575
Qy      307 gccacatcacatgaaaacggtgagcttggaaaaaagaatgctgtgtacattatgtatgc 366
Db      6576 ggccaaaaacatgaa---ggcagagctaggcaaaactgaaatgtgtgtacattatgtatgc 6632
Qy      367 caaaaagatgctgagattattatggcccaatgtcaacacaaagaagaatactaac 426
Db      6633 gaacctggcgtgaaatcattatgtacatccgcgaaacaaagaagaatctgtcgaa 6692
Qy      427 atgatagcgttgagaaatgagatgagctctgtgcgcgtgttaaaagtcaacgcggagat 486
Db      6693 atgattgaagaagtgctgtgggacaatctttaaagaagaatccagtgaanaaaggcgat 6752
Qy      487 ttttctatgtccaaagcgttactgttcatgtcgattggaagaagaatcttgcgttgag 546
Db      6753 ttttattatgtaccaaagcgttactgttcatgtcgattggaagaagaatcttattatgaa 6812
Qy      547 acgcagcagacatcagacacacacacacacacacacacacacacacacacacacacac 606
Db      6813 acgcacaaagcagtgatctactatcgtatcggtttatgtattatgacgcagacagatgaa 6872
Qy      607 ggaagctgcgcgacgttcatctgaaagaagacgtgaagtatagaggtccgtctatt 666
Db      6873 ggaagaacccggaattgacatattcaacatcatgtgtgacacacgttccgcgagaa 6932
Qy      667 ccgaacgcatcatcttaccatgaacaaattgagattggtcttaacaaagacattgatt 726

```


RESULT 6
T80389/c
ID T80389 standard; cDNA: 2111 BP.
XX T80389;
AC T80389;
XX
XX 14-APR-1998 (first entry)
DT
XX Staphylococcus aureus Gene #1 encoding cDNA sequence 1.
DE
XX Staphylococcus aureus polynucleotide and polypeptide(s) - for
XX isolating antagonist of the polypeptide(s) useful as anti-bacterials
XX
XX Claim 4: Page 26-27; 117pp; English.
PS
XX The present sequence encodes a novel polypeptide, which is optionally
XX expressed in NCIMB 40771. The polypeptide, and polynucleotide encoding
XX it, are derived from Staphylococcus aureus. Cells expressing ligands
XX binding the polypeptide can be used to isolated candidate compounds
XX that bind and inhibit the activity of the polypeptide. Such compounds
XX can be used as anti-bacterial compounds. The polypeptide may also be
XX used as an immunogen to vaccinate an animal for protection against
XX Staphylococcus aureus caused disease.
CC
XX Sequence 2111 BP; 644 A; 428 C; 385 G; 644 T; 10 other;
SQ

Query Match 23.9%; Score 225.8; DB 18; Length 2111;
Best Local Similarity 62.3%; Pred. No. 6.3e-63;
Matches 404; Conservative 0; Mismatches 238; Indels 6; Gaps 3;

OY 7 actgaacgcttaatttcaagcctgtttcaagaagaatttggcgcgagccgctta 66
DB 676 ACTATGGCATTTATTTTACAAACCGCTTTTAAAGACAGAAATTTGGGCGCATGCA 617
OY 67 gctatttgcctataccattcgcgtcaacaagaaggagagtgctgttgcgcg 126
DB 616 AAAGCATTTAATGATATCCCTAATGAACAACCTGGTGAATCTGGCAATTTTCGCA 557
OY 127 catcaaaatgttcaaacgctgttcaaaacggaatgtaagtgaggttcacgtcagagaa 186
DB 556 CATCCAATATGCTCTAATATACATTTATCATATGCTCATACAAAGATATGACCTTAGATCAA 497
OY 187 ttatgggaacatcacagacattatctgcgaagcttgaaggagagccgttcccttgc 246
DB 496 CTTTGGTCACAACTCGTGAATTTGATATGATTCAGAGATAGTTTCCCTGCTGCTA 437
OY 247 acaaaatataagatgctgcacgagactatctgcttgaagtgatgcacgaatgtgatat 306
DB 436 ACTTAAATATTAGTGCCTAATGATTAATTAATTAATTAATTAATTAATTAATTAATTA 377
OY 307 gccacaatacatgaaacggttgagcttgaaaaaagaatgctgtacattatgtattgc 366

DB 376 GCTTTAAACACGAA---GGCGAAGTGAAGTAACAGAAATGTGATAT--TTAGATGC 322
OY 367 caaaagatgctgagatatt-tatggccacaatgacacaacaagaagaactaac 425
DB 321 AGCCAGGTGCTGAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTA 262
OY 426 catgatagagctgtgaaatggagatgagcttgcgcgtgtaaggtaagccgggga 485
DB 261 CATATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTA 202
OY 486 ttcttctatggtcgaacgctgactgttcacgttgcattggaagaaggaattcgttga 545
DB 201 CTTTATTATGTTCTCTGCTGCACTGTCATGCCATTTGTCGGATTTTATTATTAGA 142
OY 546 gacgcagcagaactcagaacaactacagattatgatattgacgcgaagaatgcaga 605
DB 141 GACCAACAATCATCCATCCATGACATCCGTAATTAATGATATGATAGACAGATCCANA 82
OY 606 agcgaagctgcgagcttcatctgaaagaagcattgagtgataga 653
DB 81 TGTCAATTGCGTGAATTTGCACTTAGACGAAGTAAGATGTGATTGA 34

RESULT 7
V74477/c
ID V74477 standard; DNA: 9623 BP.
XX V74477;
XX
XX 16-MAR-1999 (first entry)
DT
XX Staphylococcus aureus contig seq ID #166.
DE
XX
XX Computer readable medium; vaccine; S. aureus infection; immunodetection;
XX cellulitis; eyelid infection; food poisoning; osteomyelitis; therapy;
XX skin infection; surgical wound infection; scalded skin syndrome; ds.
XX
XX Staphylococcus aureus.
OS
FH
FH Key Location/Qualifiers
FT misc-feature 361..420
FT /*tag= a
FT "these bases represent a line of missing text in
FT the sequence listing in the specification. They
FT are included to maintain the nucleotide numbering
FT given in the specification for this DNA sequence"
FT misc-feature 2161..2220
FT /*tag= b
FT "these bases represent a line of missing text in
FT the sequence listing in the specification. They
FT are included to maintain the nucleotide numbering
FT given in the specification for this DNA sequence"
FT misc-feature 3961..4020
FT /*tag= c
FT "these bases represent a line of missing text in
FT the sequence listing in the specification. They
FT are included to maintain the nucleotide numbering
FT given in the specification for this DNA sequence"
FT misc-feature 5761..5820
FT /*tag= d
FT "these bases represent a line of missing text in
FT the sequence listing in the specification. They
FT are included to maintain the nucleotide numbering
FT given in the specification for this DNA sequence"
FT misc-feature 7561..7620
FT /*tag= e
FT "these bases represent a line of missing text in
FT the sequence listing in the specification. They
FT are included to maintain the nucleotide numbering
FT given in the specification for this DNA sequence"
FT misc-feature 9361..9420

QY	7	acgaaccgctatcttttcaagcctgttttcaagaagaatttggcgcgacccctta	66
Db	6142	ACTATGSCATTATTTTATCAACCCGTTTTTAAAGACAGATTGGGGCGTCATCATTTA	6083
QY	67	gcgatcttggctatccacatccgcgtcaacagaaacgggaattgcttggcctttgcgcg	126
Db	6082	AAAGCATTTACATGATGATTCCTTATCGAACAACACTGGTGAAATGCTGGGCAATTTCTGCA	6023
QY	127	catcaaaatgtgtaaaagcgtgttttcaaaacgaagatgataaagggtttaaagctcaagaa	186
Db	6022	CATCCAAATGTCCTTATACATTTATTCATTTGCTCCATCAAGATATGACCTTAGATCAA	5963
QY	187	ttatgggaacatacaagacattatctcgagacagcttgaaggagaccglttccctcgtct	246
Db	5362	CTTTGGCACAAATCGTGATTTATTCGTGAATATGATTCACGAGATAGTTTCCCTCTGCTGA	5903
QY	247	acaaaaattatagtcgtgcacaggaactatcctgttccagtgatcatccgaatgatgaatat	306
Db	5502	ACTTAAATATTAAATGCGCAATGTATTAATATCTGTTCAAGTTCCACCGATGATATCATACAC	5843

the sequence listing in the specification. They

```

/NOTE="these bases represent a line of missing text in
the sequence listing in the specification. They
are included to maintain the nucleotide numbering
given in the specification for this DNA sequence"
2221..2280
/*tag= b
/NOTE="these bases represent a line of missing text in
the sequence listing in the specification. They

```

are included to maintain the nucleotide numbering given in the specification for this DNA sequence"

PN	EP786519-A2.
XX	
PD	30-JUL-1997.
XX	
PF	07-JAN-1997; 97EP-0100117.
XX	
PR	05-JAN-1996; 96US-0009861.
XX	
PA	(HUMA-) HUMAN GENOME SCI INC.

PI Barash SC,	Choi GH,	Dillon PJ,	Fannon MR,	Kunsch CA,
PI Rosen CA;				

PT Polynucleotide(s) and proteins derived from *Staphylococcus aureus*
PT stored on computer readable medium and used in the production of
PT anti-*S. aureus* vaccines

PS Claim 1; Page 255-257; 3271pp; English

CC This sequence represents one of 5191 *Staphylococcus aureus* DNA sequences
CC of the invention. The DNA sequences are recorded on a computer readable
CC medium, preferably selected from a floppy or hard disk, random access
CC memory (RAM), read-only memory (ROM), or CD-ROM. Homology searches using
CC the *S. aureus* DNA sequences allows putative functions to be assigned so
CC that protein-encoding or regulatory regions of commercial, therapeutic or
CC industrial importance can be obtained. Specifically, sequences which are
CC likely to encode antigens have been identified and these polypeptides can
CC be used in a vaccine composition against *S. aureus* infection. The
CC polypeptides can also be used in a kit for the immunodetection of
CC *S. aureus* in a sample. *S. aureus* is implicated in numerous human diseases,
CC including cellulitis, eyelid infections, food poisoning, osteomyelitis,
CC skin and surgical wound infections, scaled skin syndrome, toxic shock
CC syndrome, etc. Organisms transformed with the DNA sequences can be used
CC for recombinant production of the polypeptides. The new DNA sequences
CC (and their fragments) are useful as primers or probes for isolating
CC homologues of any of the *S. aureus* DNA sequences contained on the
CC computer readable medium.

SQ Sequence 3759 BP; 1165 A; 668 C; 510 G; 1289 T; 127 other;

Query Match	18.58;	Score 175.2;	DB 18;	Length 3759;
Best Local Similarity	47.48;	Pred. No. 2.4e-46;		
Matches 443; Conservative	1;	Mismatches 489;	Indels 1;	Gaps 1

[illegible]

OY	372	agaagccggaattatattatggcacaatacgacaacaaagaagaaactaacaccatgat	431
Dd	606	AGATGCGAAGAATGGTTATTATGGGACACTTATGCAGACGTCTAGAGAAAGAAAGTTGCCAATCATC	547
OY	432	agagcgttgagaatagtggatgagcttccttgcgccgttgtaaaagtaagccggggatltttc	491
Dd	546	TCAACAACGGAACGATATAGATTCGATCTTGATATTAATTTAAAGTAAACCCTGGGAAATTTCTA	487
OY	492	ctatgtgccaaagcgyrtaactgltcatgcatgbatbgaaaaagaatcttgccttbtgagaocga	551
Dd	486	TTTTATTNNN	427
OY	552	gcgaacctcagagcaacactacagatatatgatgatgacgcgaaaaagatgcagaaggcaa	611
Dd	426	NNNNNNNGTCAGCAATTTACTATATAGACTTTATGATTTTCAATGTCAGCAAGATATCAATATA	367
OY	612	gctgcgagatctcatctgtaaaaaagacatgaaigtatagaagytcccgtcatccaga	671
Dd	366	TGATAGACCGCTTAAATATGAAAAAGCTTTAGACGTTATTCAGTACAMTAGCACCATWTAC	307
OY	672	acgcgcatacagttccaccttagcaaatattgagatttgcttacaagaacatgatgaaty	733
Dd	306	TAAATATTTTGCCTCGAAGCGGAATTTATTGAAAACCAATASTGTACACACATTTGTTCSAA	247
OY	732	cgttaacttttcgtybgvgaatagtgaactataccaagatcacagaagcttaaagcagaacaa	791
Dd	246	TGATTTCTTACTACCTGGTTAAATGGGAATTTCTGGCACCTTAAATATATNAACCTTAG	187
OY	792	accttccttccttatcagtgibatgaaagggagggccgtatgactctcgtgtgagtgt	855
Dd	186	AGATTTCTGTTTAGTTACAGTGTGTTGGAAGCGGAAGGSCAAATGATTTGTCGATGCGTAAT	127
OY	852	ctatcccttcaaaaaaaggagacatabatgttgcgtcccttaacgylcttggagaattaaact	911
Dd	126	TTTTCAACGACACTACTGCTACAAACTTATTTTGCACCTTCGAAAGATTTGGATAGTGTCTT	67
OY	912	cgaagatagtcsgaatgatatcgtcccatctg	945
Dd	66	TGAAGGTGATTTCCATGATCATGATTAAGCTATGTC	33
RESULT	9		
V52284/c			
ID	V52284	standard; DNA; 13425 BP.	
XX	V52284;		
XX	23-OCT-1998	(first entry)	
DT			
XX	Streptococcus pneumoniae genome fragment SEQ ID NO:151.		
DE			
XX	Streptococcus pneumoniae; S. pneumoniae; genome; diagnosis; assay;		
KM	computer readable medium; vaccine; pharmaceutical composition; ds.		
RW			
XX	Streptococcus pneumoniae.		
OS			
XX	MO9818931-AZ.		
PN			
PD	07-MAY-1998.		
XX			
PF	30-OCT-1997;	97WO-US19588.	
XX			
PR	31-OCT-1996;	96US-0029960.	
XX			
PA	(HUMA-) HUMAN GENOME SCI INC.		
PI	Barash SC, Choi GH, Dillon PJ, Dougherty BA, Fannon M;		
EI	Kunsch CA, Rosen CA,		
XX	WPI; 1998-272225/24.		
DR			
CT	Computer-readable medium with recorded Streptococcus pneumoniae		

PT polynucleotide sequences - useful in diagnostic kits and assays, and
 PT pharmaceutical compositions and vaccines for Streptococcus
 pneumoniae

PS Claim 1; Page 1009-1017; 1409pp; English.

XX The present invention describes a computer readable medium which has
 CC the nucleotide sequences SEQ ID NO:1 to 391 (V52134 to V52524) recorded
 CC on it, or a representative fragment or a sequence at least 95% identical
 CC to SEQ ID NO:1 to 391. The nucleotide sequences depicted in SEQ ID NO:1
 CC to 391 (V52134 to V52524) are genomic fragments from Streptococcus
 CC pneumoniae. The present invention also describes an isolated nucleic acid
 CC molecule encoding a homologue of any of the fragments of the S.pneumoniae
 CC genome (SEQ ID NO:1 to 391) where the nucleic acid molecule is produced
 CC by a process comprising: (a) screening a genomic DNA library using as a
 CC probe a target sequence defined by any of the sequences in SEQ ID NO:1
 CC to 391, identifying members of the library which contain sequences
 CC that hybridize to the target sequence and isolating the nucleic acid
 CC molecules from the members; or (b) isolating mRNA, DNA or cDNA produced
 CC from an organism, amplifying nucleic acid molecules whose nucleotide
 CC sequence is homologous to amplification primers derived from the
 CC fragment of the S. pneumoniae genome to prime the amplification and
 CC isolating the amplified sequences. The computer readable medium can be
 CC used in a computer-based system for identifying fragments of the
 CC S. pneumoniae genome of commercial importance, or expression modulating
 CC fragments of the S. pneumoniae genome. Products from the present
 CC invention can be used in diagnosis kits and assays, and pharmaceutical
 CC compositions and vaccines for S. pneumoniae.

XX Sequence 13425 BP; 4434 A; 2362 C; 2406 G; 4223 T; 0 other;

Query Match 3.9%; Score 37.2; DB 19; Length 13425;
 Best Local Similarity 51.9%; Pred. No. 0.32; Mismatches 78; Indels 0; Gaps 0;
 Matches 84; Conservative 0;

QY 162 gtataaagggttaacgctcgaagcgaattatggaacatcacacattatcggacgct 221
 DB 7527 GAATTACGAACCTCAAAATGGAATTTATTGTAATACAAAGATATTATTACAAATT 7468
 QY 222 tgaaggagaccgttccctcctcttacaacaaatattagatgctgacgaagactatcgt 281
 DB 7467 TAAACCTAAACCTGTCATTTGCGACTACAAAGAGTAATTAATATCTGCTTTATTGG 7408
 QY 282 tcagggtcatccgaatgatgatgacacatacatgaataa 323
 DB 7407 TCATGATGATTTGAATTTAAATTTGTCAAGTACAGCAGCAAAA 7366

RESULT 10
 F22305 1082138 BP.
 ID F22305 standard; DNA: 1082138 BP.
 AC F22305;
 XX 20-MAR-2001 (first entry)
 DT Arabidopsis thaliana chromosome 4 centromere.
 XX Centromere; microsome; vector; ds.
 KM Arabidopsis thaliana.
 OS Arabidopsis thaliana.
 PM W020005325-A2.
 XX 21-SEP-2000.
 PD 17-MAR-2000; 2000WO-US07392.
 PF 18-MAR-1999; 99US-0125219.
 PR 01-APR-1999; 99US-0127409.
 PR 18-MAY-1999; 99US-0134770.
 PR 13-SEP-1999; 99US-0153584.

PR 17-SEP-1999; 99US-0154603.
 XX (UYCH-) UNIV CHICAGO.
 PA Preuss D, Copenhagen G, Keith K;
 XX WPI: 2000-587529/55.

PT Recombinant DNA construct comprising a plant centromere, useful for
 PT producing stably inherited microchromes which can serve as vectors for
 PT the construction of transgenic plant and animal cells

PS Claim 68; Page 977-1388; 1449pp; English.

XX The present invention relates to a recombinant DNA construct of a plant
 CC (Arabidopsis thaliana) centromere. The constructs are useful for
 CC producing stably inherited microchromes which can serve as vectors for
 CC the construction of transgenic plant and animal cells expressing
 CC selected proteins such as hormones, enzymes, interleukins, clotting
 CC factors, cytokines, antibodies, and growth factors.

XX Sequence 1082138 BP; 348775 A; 194404 C; 195515 G; 343444 T; 0 other;

Query Match 3.6%; Score 34.4; DB 21; Length 1082138;
 Best Local Similarity 52.0%; Pred. No. 28;
 Matches 77; Conservative 0; Mismatches 71; Indels 0; Gaps 0;

QY 259 gatgctgaccaggaactatcgttcagtgatgcacccgaatgatgataccaatcat 318
 DB 245148 ggtgatgagatgcattctattctcgtactatcttctatgaacaaagtcctt 245207
 QY 319 gaaacggtgagcttggaacaaacagaaatcgtgacatattgattgccaagaatgcc 378
 DB 245208 tctaatagaacaaaggaacaaacaaataatttaagtaatttagtccaacaaagaa 245267
 QY 379 gagattattatggtcccaatgcaacaa 406
 DB 245268 aaatctattatgagatattatgctaa 245295

RESULT 11
 T99451 6582 BP.
 ID T99451 standard; DNA: 6582 BP.
 AC T99451;
 XX 11-MAY-1998 (first entry)
 DT S. carnosus nitrate reductase molybdenum cofactor genes.
 XX Nitrite reductase molybdenum cofactor; MoEB; MoAC; MoAE;
 KM MoAD; MoBA; MoB; MoAE; MoAE; nitrate reduction; nitrite reduction;
 KM pollutant; biotransformation; water treatment; vegetable; ds.
 XX Staphylococcus carnosus DSM 10563.

XX Key Location/Qualifiers
 FH -35_signal 55..60 /*tag= a
 FT -10_signal 78..83 /*tag= b
 FT RBS 101..115 /*tag= c
 FT CDS 117..118 /*tag= d
 FT /*function= moeb gene
 FT 1150..1662 /*tag= e
 FT /*function= moab gene
 FT CDS complement (1846..2331)
 FT /*tag= f
 FT /*function= moac gene

```

FT CDS 2396..3655
FT /*tag= g
FT /function= moea gene
FT CDS 3652..4134
FT /*tag= h
FT /function= mobb gene
FT CDS 4131..4583
FT /*tag= l
FT /function= moae gene
FT CDS 4583..4816
FT /*tag= j
FT /function= moad gene
FT CDS 4821..5411
FT /*tag= k
FT /function= moba gene
FT CDS 5470..6492
FT /*tag= l
FT /function= moa gene
FT 6500..6543
FT terminator
FT /*tag= m

EP805205-A1.
05-NOV-1997.
02-MAY-1996; 96EP-0201224.
02-MAY-1996; 96EP-0201224.
02-MAY-1996; 96EP-0201224.
(NEST ) SOC PROD NESTLE SA.
Fast B, Galer W, Goeltz F, Lindgren P, Neubauer H;
Pantel I;
WPI: 1997-529059/49.
P-PSDB: W26741-49.

Recombinant S. carnosus nitrate and nitrite reductase proteins -
useful for reducing nitrate and nitrite in nitrate-polluted material

Claim 3; Page 42-49; 66pp: English.

This DNA sequence from staphylococcus carnosus includes genes that
code for the nitrate reductase molybdenum cofactor biosynthesis
proteins Moeb, Moab, Moac, Moea, Moab, Moab, Moab, Moab and Moae
(see W26741-49). This gene region was identified following Tn917
insertion mutagenesis studies. Claimed DNA molecules (see
T99449-51) can be used in the recombinant production of claimed
proteins (see W26733-49) involved in nitrate and nitrite reduction
in S. carnosus. These recombinant proteins, or cells expressing
them, in free or immobilised form, can be used in a claimed method
for reducing nitrate and nitrite in nitrate-polluted water,
vegetables and other polluted materials.

Sequence 6582 BP; 2286 A; 980 C; 1343 G; 1973 T; 0 other;

```

Query Match 3.6%; Score 33.8; DB 18; Length 6582;
 Best Local Similarity 49.2%; Pred. No. 2.7; Mismatches 92; Indels 0; Gaps 0;
 Matches 89; Conservative 0;

```

OY 233 gttccctcgtctcaaaaaatattagatgctgacagactattcgttcagtgatc 292
    ||||| || || || || || || || || || || || || || || || || ||
DB 5248 gttgcgcatattatgcacgctgcttagatgcagatgactatgtagtcgttattc 5307
    || || || || || || || || || || || || || || || || || || ||
OY 293 cgaatgtagatgacacatatcagaacagctgagcttggaaaaacaatgctggt 352
    || || || || || || || || || || || || || || || || || || ||
DB 5308 aacaaacgcgtctcgtatgtagatgtagatgtagtgcacgattcacaagaatggtata 5367
    || || || || || || || || || || || || || || || || || || ||
OY 333 acattatgattgccaagaagatgcgagatattattatgccaacatgcacaacaagaag 412
    || || || || || || || || || || || || || || || || || || ||
DB 5368 agaataattatccccaagaatttagagagacataaagaataatacaacagtttaacy 5427
    || || || || || || || || || || || || || || || || || || ||

```

OY 413 a 413
 DB 5428 a 5428

```

RESULT 12
ID Q15172 standard; cDNA; 3475 BP.
XX Q15172;
AC Q15172;
DE 08-MAR-1992 (first entry)
XX Human erythrocyte membrane anion-transport protein (band 3) gene..
XX Senescent cell peptide; ANION 1; COOH; auto immune disease.
XX Homo sapiens.
XX Key Location/Qualifiers
FT CDS 151..2886
FT /*tag= a

WO9118013-A.
PD 28-NOV-1991.
XX 24-MAY-1991; 91WO-US03557.
PF 24-MAY-1991; 91WO-US03557.
XX 24-MAY-1990; 90US-0528210.
PR (RESE ) RES CORP TECH INC.
XX (RESE ) RES CORP TECH INC.
PA Kay MBM, Marchalonis JJ;
XX WPI: 1991-369184/50.
DR P-PSDB: R15355.
XX Synthetic senescent cell antigen and peptide(s) - used to detect
PT antibodies to SCA, study cellular ageing and auto-immune
PT mechanisms, separate anions and treat auto-immune disease
XX
XX Disclosure: Fig 1; 74pp: English.
XX The sequence is that of the cDNA encoding the human erythrocyte
XX anion-transport protein (band 3). Nucleotides 1612-1662 (ANION 1) or
XX 2434-2481 (COOH) of this sequence encode a peptide which is immuno-
XX reactive with antibodies to senescent cell antigen (SCA). The antigen
XX and peptides can be used in any method or technique involving or
XX requiring the inhibition of the binding of senescent cell IgG to
XX senescent cells in vitro or in vivo. The prods. can be used in to
XX examine various physiological, biochemical, and immunological
XX mechanisms in humans and other animals, such as those relating to
XX cellular ageing, degeneration, and life span; removal of senescent
XX cells; anion transport across cellular membranes; disease
XX mechanisms where increased senescent cell IgG may play a role, such
XX as haemolytic anaemia, sickle cell anaemia, and idiopathic
XX thrombocytopenia purpura; other autoimmune disease mechanisms, such
XX as rheumatoid arthritis and systemic lupus erythematosus; and other
XX diseases such as those involved in malaria and cancer. More
XX specifically, the prods. are useful in cosmetic and pharmaceutical
XX compsns., diagnostic kits, and methods for detecting or measuring
XX Abs to SCA, studying cellular ageing and autoimmune mechanisms,
XX separating anions from a gas or liquid, or diagnosis and treatment
XX of certain diseases.
XX
XX Sequence 3475 BP; 672 A; 1050 C; 997 G; 756 T; 0 other;

```

Query Match 3.5%; Score 33.2; DB 12; Length 3475;
 Best Local Similarity 50.6%; Pred. No. 3;
 Matches 80; Conservative 0; Mismatches 78; Indels 0; Gaps 0;

•
•
•
•

GenCore version 4.5
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: May 17, 2001, 14:52:15 ; Search time 67.19 Seconds
(without alignments)
2455.716 Million cell updates/sec

Title: US-09-462-846-1

Perfect score: 945
Sequence: 1 atgaagcactgaaccgttatt.....aatgtatcgttcacatctg 945

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 302621 segs, 87301344 residues

Total number of hits satisfying chosen parameters: 605242

Minimum DB seq length: 0
Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : Issued Patents, NA: *
1: /cgn2_6/ptodata/1/ina/5A.COMB.seq: *
2: /cgn2_6/ptodata/1/ina/5B.COMB.seq: *
3: /cgn2_6/ptodata/1/ina/6A.COMB.seq: *
4: /cgn2_6/ptodata/1/ina/6B.COMB.seq: *
5: /cgn2_6/ptodata/1/ina/PCNUS.COMB.seq: *
6: /cgn2_6/ptodata/1/ina/backfiles1.seq: *

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Query Match	Length	ID	Description
C 1	50	5.3	7218	1	US-08-232-463-14
C 2	32.2	3.4	495	4	US-08-964-268-19
C 3	30.6	3.2	3600	3	US-08-894-731-1
C 4	30.2	3.2	10299	2	US-08-477-451-1
C 5	30.2	3.2	10299	2	US-08-477-451-5
C 6	30.2	3.2	19932	2	US-08-477-451-25
C 7	29.8	3.2	3128	4	US-08-716-449-1
C 8	29.6	3.1	2185	4	US-08-714-918-83
C 9	29.6	3.1	2185	4	US-09-265-315-83
C 10	29.6	3.1	2185	4	US-09-265-315-83
C 11	29.2	3.1	1549	1	US-08-317-880-5
C 12	29.2	3.1	1549	2	US-08-783-396-5
C 13	29.2	3.1	2471	1	US-08-152-271-1
C 14	29.2	3.1	3954	1	US-08-317-880-1
C 15	29.2	3.1	3954	1	US-08-782-396-1
C 16	28.8	3.0	1448	1	US-08-585-758A-1
C 17	28.8	3.0	1448	1	US-08-977-818-1
C 18	28.8	3.0	1448	2	US-08-670-274B-1
C 19	28.8	3.0	12720	1	US-08-403-866-11
C 20	28.8	3.0	16442	3	US-08-781-891-208
C 21	28.6	3.0	1829	2	US-08-687-080-57
C 22	28.6	3.0	2214	6	5258502-1
C 23	28.4	3.0	2256	2	US-08-896-590A-3
C 24	28.4	3.0	2259	2	US-08-896-590A-1
C 25	28.4	3.0	12047	2	US-09-022-461-1
C 26	28.2	3.0	1244	2	US-08-591-685-12
C 27	28.2	3.0	2479	4	US-09-050-863-4

C 28	28.2	3.0	5022	2	US-09-289-368-1	Sequence 1, Appl
C 29	28.2	3.0	5183	2	US-09-039-555B-18	Sequence 18, Appl
C 30	28.2	3.0	5243	3	US-08-414-335-2	Sequence 2, Appl
C 31	28	3.0	1209	4	US-08-822-774-18	Sequence 16, Appl
C 32	28	3.0	7400	4	US-07-674-852-1	Sequence 1, Appl
C 33	28	3.0	7400	4	US-08-473-185-1	Sequence 1, Appl
C 34	28	3.0	9171	1	US-08-038-682-5	Sequence 5, Appl
C 35	28	3.0	9171	1	US-08-302-832-5	Sequence 5, Appl
C 36	28	3.0	9171	2	US-08-530-198-5	Sequence 5, Appl
C 37	28	3.0	9171	2	US-08-469-880-5	Sequence 5, Appl
C 38	28	3.0	9171	2	US-08-728-470-5	Sequence 5, Appl
C 39	28	3.0	9171	2	US-08-617-697-5	Sequence 5, Appl
C 40	28	3.0	9323	1	US-08-038-682-6	Sequence 6, Appl
C 41	28	3.0	9323	1	US-08-302-832-6	Sequence 6, Appl
C 42	28	3.0	9323	2	US-08-530-198-6	Sequence 6, Appl
C 43	28	3.0	9323	2	US-08-469-880-6	Sequence 6, Appl
C 44	28	3.0	9323	2	US-08-728-470-6	Sequence 6, Appl
C 45	28	3.0	9323	2	US-08-617-697-6	Sequence 6, Appl

ALIGNMENTS

RESULT 1
US-08-232-463-14/c
; Sequence 14, Application US/08232463
; Patent No. 5670367
; GENERAL INFORMATION:
; APPLICANT: DORNER, F.
; APPLICANT: SCHEIFLINGER, F.
; APPLICANT: FALKNER, F. G.
; TITLE OF INVENTION: RECOMBINANT FOWLPOX VIRUS
; NUMBER OF SEQUENCES: 52
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Foley & Lardner
; STREET: 1800 Diagonal Road, Suite 500
; CITY: Alexandria
; STATE: VA
; COUNTRY: USA
; ZIP: 22313-0299
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patentin Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/232,463
; FILING DATE:
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US/07/935,313
; FILING DATE:
; APPLICATION NUMBER: EP 91 114 300.6
; FILING DATE: 26-AUG-1991
; ATTORNEY/AGENT INFORMATION:
; NAME: BENT, Stephen A.
; REGISTRATION NUMBER: 29,768
; REFERENCE/DOCKET NUMBER: 30472/114 IMMU
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (703)836-9300
; TELEFAX: (703)863-4109
; TELEX: 899149
; INFORMATION FOR SEQ ID NO: 14:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 7218 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; IMMEDIATE SOURCE:
; CLONE: pTZgpt-Fls
; US-08-232-463-14

```

Query Match      5.3%; Score 50; DB 1; Length 7218;
Best Local Similarity 3.0%; Pred. No. 5e-06;
Matches 11; Conservative 208; Mismatches 143; Indels 0; Gaps 0;

QY 294 gaatgtagatgccacacatcatgaaacggtgacttggaacacagaatgctgtgta 353
    || || || || || | : : : : : : : : : : : : : : : : : : : : : :
Db 1452 GATAGAAAGATTGTGACRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRR 1393

QY 354 catatgattgccaacacatgcagatatattatggccaacacacacacacaga 413
    : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 1392 RRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRR 1333

QY 414 agaacactacacatcatgtagcgtggaatggaatgagctctgcgcgtgtgaagt 473
    : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 1332 RRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRR 1273

QY 474 aagccgggggatttttctatgtgccaagcggtactgtctcgtgatggaaagat 533
    : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 1272 RRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRR 1213

QY 534 tcttgcttgagacgcagacagacacacacacacacacacacacacacacacac 593
    : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 1212 RRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRR 1153

QY 594 aaaagatgcagaagcagacgtcgcgagctcatctgaaacacacacacacacacac 653
    : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 1152 RRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRR 1093

QY 654 gg 655
    : :
Db 1092 RR 1091

RESULT 2
US-08-964-268-19/c
; Sequence 19, Application US/08964268
; Patent No. 6114503
; GENERAL INFORMATION:
; APPLICANT: WEI, YING-BEI
; APPLICANT: RUBEN, STEVEN M
; APPLICANT: SANCAR, AZIZ
; APPLICANT: HSU, SHIAO-WEN D
; APPLICANT: KAZANTSEV, ALEKSEY G
; TITLE OF INVENTION: HUMAN BLUE-LIGHT PHOTORECEPTOR hCRY2
; NUMBER OF SEQUENCES: 22
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: STERNE, KESSLER, GOLDSTEIN & FOX P.L.L.C.
; STREET: 1100 NEW YORK AVENUE, NW, SUITE 600
; CITY: WASHINGTON
; STATE: DC
; COUNTRY: USA
; ZIP: 20005-3934
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patentin Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/964,268
; FILING DATE: HERewith
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 60/030,189
; FILING DATE: 04-NOV-1996
; ATTORNEY/AGENT INFORMATION:
; NAME: STEFFE, ERIC K.
; REGISTRATION NUMBER: 36,688
; REFERENCE/DOCKET NUMBER: 1488.0750001/EKS
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 202-371-2600
; TELEFAX: 202-371-2540
; INFORMATION FOR SEQ ID NO: 19:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 495 base pairs

```

```

; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: cDNA
US-08-964-268-19

Query Match      3.4%; Score 32.2; DB 4; Length 495;
Best Local Similarity 47.3%; Pred. No. 0.64;
Matches 61; Conservative 0; Mismatches 68; Indels 0; Gaps 0;

QY 318 tgaacacggtgagcttggaacacagaatgctgtgacattatgattgccaacagatgc 377
    || || || || || | : : : : : : : : : : : : : : : : : : : : : :
Db 426 TGAANAAGTTGAACNGGMAANGNNNTGCCATGCCCTAAGNNNTGGAATAATTAATTC 367

QY 378 cgagattattatgcccacacatgcacacacacacacacacacacacacacacacac 437
    || || || || || | : : : : : : : : : : : : : : : : : : : : : :
Db 366 CGAAGTTGTCNTGCCAACNTNTAACCTNAGGGAAGAAAGAACNCCANACAGAGNC 307

QY 438 tggagaatg 446
    || || || |
Db 306 AGGAAAAG 298

RESULT 3
US-08-894-731-1
; Sequence 1, Application US/08894731
; Patent No. 6084089
; GENERAL INFORMATION:
; APPLICANT: MINE, TOSHIKI
; APPLICANT: OHYAMA, AKIO
; APPLICANT: HIYOSHI, TORU
; APPLICANT: KASAKURA, KEISUKE
; TITLE OF INVENTION: COLD-INDUCIBLE PROMOTER SEQUENCE
; FILE REFERENCE: 760-234P
; CURRENT APPLICATION NUMBER: US/08/894,731
; CURRENT FILING DATE: 1997-10-27
; NUMBER OF SEQ ID NOS: 8
; SOFTWARE: Patentin Ver. 2.0
; SEQ ID NO 1
; LENGTH: 3600
; TYPE: DNA
; ORGANISM: Solanum tuberosum
US-08-894-731-1

Query Match      3.2%; Score 30.6; DB 3; Length 3600;
Best Local Similarity 49.7%; Pred. No. 6.5;
Matches 78; Conservative 0; Mismatches 79; Indels 0; Gaps 0;

QY 216 acagctgaaggagacgttccctctgcttacaacacacacacacacacacacacacac 275
    || || || || || | : : : : : : : : : : : : : : : : : : : : : :
Db 156 aaaattgtgttactattatttatttatttatttatttatttatttatttatttattt 215

QY 276 atctgtcaggtgcatccgaatgataatgacacacacacacacacacacacacacacac 335
    || || || || || | : : : : : : : : : : : : : : : : : : : : : :
Db 216 aaatgggttgacatcactactaaataatgaaaacacacacacacacacacacacac 275

QY 336 aaaaacagaatgctgtacattatgattgccaanaa 372
    || | || || || | : || || || || || || || || || || || || || || ||
Db 276 aaattgacaatgttcacatgattttttttcatgcaaa 312

RESULT 4
US-08-477-451-1
; Sequence 1, Application US/08477451
; Patent No. 592865
; GENERAL INFORMATION:
; APPLICANT: Covacci, Antonello
; TITLE OF INVENTION: Helicobacter Pylori CagI Region
; NUMBER OF SEQUENCES: 46
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Chiron Corporation

```


STREET: 4560 Horton Street
CITY: Emeryville
STATE: CA
COUNTRY: USA
ZIP: 94608-2916
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/477,451
FILING DATE: 07-JUN-1995
CLASSIFICATION: 435
ATTORNEY/AGENT INFORMATION:
NAME: McClung, Barbara G.
REGISTRATION NUMBER: 33,113
REFERENCE/DOCKET NUMBER: 0335.002
TELECOMMUNICATION INFORMATION:
TELEPHONE: 510-601-2708
TELEFAX: 510-655-3542
INFORMATION FOR SEQ ID NO: 1:
SEQUENCE CHARACTERISTICS:
LENGTH: 10299 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: DNA (genomic)
US-08-477-451-1

Query Match 3.28; Score 30.2; DB 2; Length 10299;
Best Local Similarity 52.8%; Pred. No. 16;
Matches 65; Conservative 0; Mismatches 58; Indels 0; Gaps 0;
QY 740 ttctggtggaatggaactatcagatcagcaagcttaagcagcaaaaccattcc 799
DB 4195 TTTTGCTCGGCACTGACATCATTCAGAGATGATGAACCTACTGTCACACCATTTTTC 4254
QY 800 ttctatcagtgatgtaaggaggcgatgatctctgtgagatgctatcctt 859
DB 4255 TTTCTCAAGTCGTGTAGAAATTGTAGATCTTTAGATCTTTGATGATTTTAAATCGTC 4314
QY 860 tca 862
DB 4315 TCA 4317

RESULT 5
US-08-477-451-5/c
Sequence 5, Application US/08477451
Patent No. 5928865
GENERAL INFORMATION:
APPLICANT: Covacci, Antonello
TITLE OF INVENTION: Helicobacter Pylori CagI Region
NUMBER OF SEQUENCES: 46
CORRESPONDENCE ADDRESS:
ADDRESSEE: Chiron Corporation
STREET: 4560 Horton Street
CITY: Emeryville
STATE: CA
COUNTRY: USA
ZIP: 94608-2916
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/477,451
FILING DATE: 07-JUN-1995
CLASSIFICATION: 435
ATTORNEY/AGENT INFORMATION:

NAME: McClung, Barbara G.
REGISTRATION NUMBER: 33,113
REFERENCE/DOCKET NUMBER: 0335.002
TELECOMMUNICATION INFORMATION:
TELEPHONE: 510-601-2708
TELEFAX: 510-655-3542
INFORMATION FOR SEQ ID NO: 5:
SEQUENCE CHARACTERISTICS:
LENGTH: 10299 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: DNA (genomic)
US-08-477-451-5

Query Match 3.28; Score 30.2; DB 2; Length 10299;
Best Local Similarity 52.8%; Pred. No. 16;
Matches 65; Conservative 0; Mismatches 58; Indels 0; Gaps 0;
QY 740 ttctggtggaatggaactatcagatcagcaagcttaagcagcaaaaccattcc 799
DB 6105 TTTTGCTCGGCACTGACATCATTCAGAGATGATGAACCTACTGTCACACCATTTTTC 6046
QY 800 ttctatcagtgatgtaaggaggcgatgatctctgtgagatgctatcctt 859
DB 6045 TTTCTCAAGTCGTGTAGAAATTGTAGATCTTTAGATCTTTGATGATTTTAAATCGTC 5986
QY 860 tca 862
DB 5985 TCA 5983

RESULT 6
US-08-477-451-25
Sequence 25, Application US/08477451
Patent No. 5928865
GENERAL INFORMATION:
APPLICANT: Covacci, Antonello
TITLE OF INVENTION: Helicobacter Pylori CagI Region
NUMBER OF SEQUENCES: 46
CORRESPONDENCE ADDRESS:
ADDRESSEE: Chiron Corporation
STREET: 4560 Horton Street
CITY: Emeryville
STATE: CA
COUNTRY: USA
ZIP: 94608-2916
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/477,451
FILING DATE: 07-JUN-1995
CLASSIFICATION: 435
ATTORNEY/AGENT INFORMATION:
NAME: McClung, Barbara G.
REGISTRATION NUMBER: 33,113
REFERENCE/DOCKET NUMBER: 0335.002
TELECOMMUNICATION INFORMATION:
TELEPHONE: 510-601-2708
TELEFAX: 510-655-3542
INFORMATION FOR SEQ ID NO: 25:
SEQUENCE CHARACTERISTICS:
LENGTH: 19932 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: DNA (genomic)
US-08-477-451-25

```

? ORGANELLLE:
? IMMEDIATE SOURCE:
? LIBRARY: lambda Zap tuber cDNA
? CLONE: 1.2.1 and E2
? POSITION IN GENOME:
? CHROMOSOME/SEGMENT:
? MAP POSITION:
? UNITS:
? FEATURE: open reading frame
? NAME/KEY: starch branching enzyme
? LOCATION: 44-2788
? IDENTIFICATION METHOD: lone ORF with homology to other starch
? IDENTIFICATION METHOD: branching enzymes
? OTHER INFORMATION: complements KV832 E coli glycogen
? OTHER INFORMATION: branching enzyme mutant
?
? US-08-716-449-1

```

[illegible]

TELEX: 67-3510

```
; INFORMATION FOR SEQ ID NO: 83:
; SEQUENCE CHARACTERISTICS:
```

```

; INFORMATION FOR SEQ ID NO: 83
; SEQUENCE CHARACTERISTICS:
;   LENGTH: 2185 base pairs

```

RESULT 13
US-08-152-271-1
; Sequence 1, Application US/08152271
; Patent No. 5583039
; GENERAL INFORMATION:

APPLICANT: Park, Kwan Hwa
 APPLICANT: Kim, In Cheol
 APPLICANT: Kim, Key Hyup
 APPLICANT: Cha, Jae Ho
 APPLICANT: Jang, So Young
 APPLICANT: Kim, Jeong Ryel
 APPLICANT: Seo, Byung Cheol
 APPLICANT: Choi, Yang Do
 TITLE OF INVENTION: A Method for Isolating Gene of
 TITLE OF INVENTION: Maltogenic Amylase and to Produce
 TITLE OF INVENTION: the Enzyme from the Gene.
 NUMBER OF SEQUENCES: 1
 CORRESPONDENCE ADDRESS:
 ADDRESSEE: Kwan Hwa Park
 ADDRESSEE: Department of Food Science and Technology,
 ADDRESSEE: College of Agriculture and Life,
 ADDRESSEE: Seoul National University
 STREET: 103 Seodoo-dong Kwensun-ku
 CITY: Suwon
 STATE: Kyonggi-Do
 COUNTRY: Korea
 ZIP: 441-744
 COMPUTER READABLE FORM:
 MEDIUM TYPE: 5.25 inch diskette- 360 Kb storage
 COMPUTER: IBM PC
 OPERATING SYSTEM: MS-DOS
 SOFTWARE: Wordperfect 5.1 (stored as an ASCII file)
 CURRENT APPLICATION DATA:
 APPLICATION NUMBER: US/08/152,271
 FILING DATE:
 CLASSIFICATION: 435
 PRIOR APPLICATION DATA:
 APPLICATION NUMBER: 07/ 534,679
 FILING DATE: 07-JUN-1990
 APPLICATION NUMBER: KOREAN No. 5583039 89-7818
 FILING DATE: 07-JUN-1989
 ATTORNEY/AGENT INFORMATION:
 NAME: DUNN, MICHAEL L.
 REGISTRATION NUMBER: 25,330
 REFERENCE/DOCKET NUMBER: H2P:10405
 TELECOMMUNICATION INFORMATION:
 TELEPHONE: ATTORNEY: 716-433-1661
 TELEFAX: ATTORNEY: 716-433-1665
 INFORMATION FOR SEQ ID NO: 1:
 SEQUENCE CHARACTERISTICS:
 LENGTH: 2471 nucleotides
 TYPE: Nucleic acids
 STRANDEDNESS: double stranded
 TOPOLOGY: circular
 MOLECULE TYPE: Genomic DNA
 DESCRIPTION: The nucleotide sequence of 2471 bp DNA frag-
 DESCRIPTION: ment of B. licheniformis containing an open reading frame
 DESCRIPTION: capable of encoding a maltogenic amylase, its 5' and 3'
 DESCRIPTION: sequences, and 578 amino acid deduced sequence of the
 DESCRIPTION: open reading frame.
 HYPOTHEICAL: NO
 ANTI-SENSE: NO
 FRAGMENT TYPE: contains N-, C-, and internal fragments.
 ORIGINAL SOURCE:
 ORGANISM: Gram positive Bacterium
 STRAIN: Bacillus licheniformis
 INDIVIDUAL ISOLATE: ATCC 27811
 DEVELOPMENTAL STAGE: vegetative cells
 IMMEDIATE SOURCE:
 CLONE: p10322
 FEATURE:
 NAME/KEY: B. licheniformis Maltogenic Amylase
 LOCATION: putative promoter; 197-202 (-35), 235-240 (-10)
 LOCATION: putative ribosome binding site; 282-287
 LOCATION: putative translation initiation site; 292
 LOCATION: translation termination site; 2076
 IDENTIFICATION METHOD: by similarity with known sequence
 IDENTIFICATION METHOD: or to an established consensus

IDENTIFICATION METHOD: sequence
 OTHER INFORMATION: The E. coli transformants harboring
 OTHER INFORMATION: the clone containing the sequence showed starch hydrolyz-in
 OTHER INFORMATION: phenocopy on a starch agar plate only after the cell
 OTHER INFORMATION: membrane was disrupted with D-cycloserine. The gel
 OTHER INFORMATION: product hydrolyzes starch, pullulan, and cyclodextrin and
 OTHER INFORMATION: has transferase activity.
 PUBLICATION INFORMATION:
 AUTHORS: KIM, IN CHEOL
 AUTHORS: CHA, JAE HO
 AUTHORS: KIM, JEONG RYUL
 AUTHORS: JANG, SO YOUNG
 AUTHORS: SEO, BYUNG CHEOL
 AUTHORS: CHEONG, TAE KYOU
 AUTHORS: LEE, DAE SIL
 AUTHORS: CHOI, YANG DO
 AUTHORS: PARK, KWAN HWA
 TITLE: CATALYTIC PROPERTIES OF THE CLONED AMYLASE FROM
 TITLE: BACILLUS licheniformis.
 JOURNAL: THE JOURNAL OF BIOLOGICAL CHEMISTRY
 VOLUME: 267
 ISSUE: NO. 31
 PAGES: 22108-22114
 DATE: 05-NOV-1992
 RELEVANT RESIDUES IN SEQ ID NO: 1 : from 1 bp to 2471
 US-08-152-271-1

Query Match 3.1%; Score 29.2; DB 1; Length 2471;
 Best Local Similarity 48.8%; Pred. No. 15;
 Matches 79; Conservative 0; Mismatches 83; Indels 0; Gaps 0;

Oy 735 ttacttcggtgggaatggaactatcacgagtcagcaagctaaagcaaaacc 794
 Db 1997 tttcatttttggggggtctaaattctgttagcgcttaacagctttaaacattgttagc 2056
 Oy 795 attccttctacagtgatgaaagggagggcgatagatctctgtgagatgctta 854
 Db 2057 gttttgtgcagagattacattgtctcagattccaagatgaaggggaaaaaagatgca 2116
 Oy 855 tcccttcaaaaagagagatcatatgttcctccttaaggtct 896
 Db 2117 acattttaaaaaaagaaatttttagtctgtcgccgtcactgtct 2158

RESULT 14
 US-08-317-880-1
 ; Sequence 1, Application US/08317880
 ; Patent No. 5593862
 ; GENERAL INFORMATION:
 APPLICANT: Hall, Linda M.
 TITLE OF INVENTION: Proteins for Cation Channel Expression
 TITLE OF INVENTION: and Function
 NUMBER OF SEQUENCES: 5
 CORRESPONDENCE ADDRESS:
 ADDRESSEE: Nixon, Hargrave, Devans & Doyle
 STREET: Clinton Square, P.O. Box 1051
 CITY: Rochester
 STATE: New York
 COUNTRY: U.S.A.
 ZIP: 14603
 COMPUTER READABLE FORM:
 MEDIUM TYPE: Floppy disk
 COMPUTER: IBM PC compatible
 OPERATING SYSTEM: PC-DOS/MS-DOS
 SOFTWARE: Patentin Release #1.0, Version #1.30
 CURRENT APPLICATION DATA:
 APPLICATION NUMBER: US/08/317,880
 FILING DATE: 04-OCT-1994
 CLASSIFICATION: 435
 ATTORNEY/AGENT INFORMATION:
 NAME: Rogalsky, Peter
 REGISTRATION NUMBER: 38,601

Fri May 18 10:37:59 2001

us-09-462-846-1.rml

Page 9

GenCore version 4.5
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: May 17, 2001, 11:56:34 ; Search time 1125.86 Seconds
(without alignments)
7332.717 Million cell updates/sec

Title: US-09-462-846-1

Perfect score: 945
Sequence: 1 atgacgactgaacgcttatt.....aatgtatcgtctcccatctg 945

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 9623517 seqs, 4368049070 residues

Total number of hits satisfying chosen parameters: 19247034

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database :

EST:*

1: gb_est1:*
2: gb_est2:*
3: gb_est3:*
4: gb_est4:*
5: gb_est5:*
6: gb_est6:*
7: gb_est7:*
8: gb_est8:*
9: gb_est9:*
10: gb_est10:*
11: gb_est11:*
12: gb_est12:*
13: gb_est13:*
14: gb_est14:*
15: gb_est15:*
16: gb_est16:*
17: gb_est17:*
18: gb_est18:*
19: gb_est19:*
20: gb_est20:*
21: gb_est21:*
22: gb_est22:*
23: gb_est23:*
24: gb_est24:*
25: gb_est25:*
26: gb_est26:*
27: gb_est27:*
28: gb_est28:*
29: gb_est29:*
30: gb_est30:*
31: gb_est31:*
32: gb_est32:*
33: gb_est33:*
34: gb_est34:*
35: gb_est35:*
36: gb_est36:*
37: gb_est37:*
38: gb_est38:*
39: gb_est39:*
40: gb_est40:*
41: gb_est41:*
42: gb_est42:*
43: gb_est43:*

44: em_esthum10:*
45: em_esthum11:*
46: em_esthum12:*
47: em_esthum13:*
48: em_esthum14:*
49: em_esthum15:*
50: em_esthum16:*
51: em_esthum17:*
52: em_esthum18:*
53: em_esthum19:*
54: em_esthum20:*
55: em_esthum21:*
56: em_esthum22:*
57: em_esthum23:*
58: em_esthum24:*
59: em_esthum25:*
60: em_esthum26:*
61: em_esthum27:*
62: em_esthum28:*
63: em_esthum29:*
64: em_esthum30:*
65: em_esthum31:*
66: em_esthum32:*
67: em_esthum33:*
68: em_esthum34:*
69: em_esthum35:*
70: em_esthum36:*
71: em_esthum37:*
72: em_esthum38:*
73: em_esthum39:*
74: em_esthum40:*
75: em_esthum41:*
76: em_esthum42:*
77: em_esthum43:*
78: em_esthum44:*
79: em_esthum45:*
80: em_esthum46:*
81: em_esthum47:*
82: em_esthum48:*
83: em_esthum49:*
84: em_esthum50:*
85: em_esthum51:*
86: em_esthum52:*
87: em_esthum53:*
88: em_esthum54:*
89: em_esthum55:*
90: em_esthum56:*
91: em_esthum57:*
92: em_esthum58:*
93: em_esthum59:*
94: em_esthum60:*
95: em_esthum61:*
96: em_esthum62:*
97: em_esthum63:*
98: em_esthum64:*
99: em_esthum65:*
100: em_esthum66:*
101: em_esthum67:*
102: em_esthum68:*
103: em_esthum69:*
104: em_esthum70:*
105: em_esthum71:*
106: em_esthum72:*
107: em_esthum73:*
108: em_esthum74:*
109: em_esthum75:*
110: em_esthum76:*
111: em_esthum77:*
112: em_esthum78:*
113: em_esthum79:*
114: em_esthum80:*
115: em_esthum81:*
116: em_esthum82:*

117: gb_est48:*
 118: gb_est49:*
 119: gb_est50:*
 120: gb_est51:*
 121: gb_est52:*
 122: gb_est53:*
 123: gb_est54:*
 124: gb_est55:*
 125: gb_est56:*
 126: gb_est57:*
 127: gb_est58:*
 128: gb_est59:*
 129: gb_est60:*
 130: gb_est61:*
 131: gb_est62:*
 132: gb_est63:*
 133: gb_est64:*
 134: gb_est65:*
 135: gb_est66:*
 136: gb_est67:*
 137: gb_est68:*
 138: gb_est69:*
 139: gb_est70:*
 140: gb_est71:*
 141: gb_est72:*
 142: gb_est73:*
 143: gb_est74:*
 144: gb_est75:*
 145: gb_est76:*
 146: gb_est77:*
 147: gb_est78:*
 148: gb_est79:*
 149: gb_est80:*
 150: gb_est81:*
 151: gb_est82:*
 152: gb_est83:*
 153: gb_est84:*
 154: gb_est85:*
 155: gb_est86:*
 156: gb_est87:*
 157: gb_est88:*
 158: gb_est89:*
 159: gb_est90:*
 160: gb_est91:*
 161: gb_est92:*
 162: gb_est93:*
 163: gb_est94:*
 164: gb_est95:*
 165: gb_est96:*
 166: gb_est97:*
 167: gb_est98:*
 168: gb_est99:*
 169: gb_est100:*
 170: gb_est101:*
 171: gb_est102:*
 172: gb_est103:*
 173: gb_est104:*
 174: gb_est105:*
 175: gb_est106:*
 176: gb_est107:*
 177: gb_est108:*
 178: gb_est109:*
 179: gb_est110:*
 180: gb_est111:*
 181: gb_est112:*
 182: gb_est113:*
 183: gb_est114:*
 184: gb_est115:*
 185: gb_est116:*
 186: gb_est117:*
 187: gb_est118:*
 188: gb_est119:*
 189: gb_est120:*

190: em_gss_pin1:*
 191: em_gss_pin2:*
 192: em_gss_pro:*
 193: em_gss_rod1:*
 194: em_gss_rod2:*
 195: em_gss_rod3:*
 196: em_gss_rod4:*
 197: em_gss_rod5:*
 198: em_gss_rod6:*
 199: em_gss_vrt1:*
 200: em_gss_vrt2:*
 201: em_gss_vrt3:*
 202: em_gss_vrt4:*
 203: em_gss_vrt5:*
 204: em_gss_vrt6:*
 205: em_gss_vrt7:*
 206: em_gss_vrt8:*
 207: em_gss_vrt9:*
 208: em_gss_vrt10:*
 209: em_gss_vrt11:*
 210: em_gss_vrt12:*
 211: em_gss_vrt13:*
 212: em_gss_vrt14:*
 213: em_gss_vrt15:*
 214: em_gss_vrt16:*
 215: em_gss_vrt17:*
 216: em_gss_vrt18:*
 217: em_gss_vrt19:*
 218: em_gss_vrt20:*
 219: em_gss_vrt21:*
 220: em_gss_vrt22:*
 221: em_gss_vrt23:*
 222: em_gss_vrt24:*
 223: em_gss_vrt25:*
 224: em_gss_vrt26:*
 225: em_gss_vrt27:*
 226: em_gss_vrt28:*
 227: em_gss_vrt29:*
 228: em_gss_vrt30:*
 229: em_gss_vrt31:*
 230: em_gss_vrt32:*
 231: em_gss_vrt33:*
 232: em_gss_vrt34:*
 233: em_gss_vrt35:*
 234: em_gss_vrt36:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
C 1	48	5.1	997	CNS0057E	AL060767 Drosophila
C 2	41.2	4.4	1204	CNS016E2	AL106628 Drosophila
C 3	39.8	4.2	292	AO070844	AO070844 HS-2256_A
C 4	38.6	4.1	1095	CNS05PY4	AL348529 Tetradon
C 5	38.4	4.1	419	AO675159	AO675159 HS-2162_A
C 6	38.2	4.0	395	AA550191	AA550191 1307m3 gm
C 7	38.2	4.0	712	BF628576	BF628576 HVSMB000
C 8	38.2	4.0	375	AO137339	AO137339 HS-3058_B
C 9	37.8	4.0	635	BF815061	BF815061 MR2-C1012
C 10	37.6	4.0	889	AZ209683	AZ209683 SP-0138_A
C 11	37.6	3.9	799	AQ842470	AQ842470 CP61126A
C 12	37	3.9	1079	CNS02H03	AL197004 Tetradon
C 13	36.8	3.9	411	BF727776	BF727776 1000052F0
C 14	36.8	3.9	1101	CNS006DR	AL072060 Drosophila
C 15	36.6	3.9	229	CNS0039G	AL063921 Drosophila
C 16	36.4	3.9	249	AV231078	AV231078 AV231078
C 17	36.4	3.9	966	CNS00154	AL075573 Drosophila
C 18	35.8	3.8	306	BB200489	BB200489 BB200489

```

c 19 35.8 3.8 451 15 A1071988
c 20 35.8 3.8 855 229 CNS018PE
c 21 35.8 3.8 1101 229 CNS0182P
c 22 35.6 3.8 504 115 AW463593
c 23 35.6 3.8 530 201 A0075795
c 24 35.4 3.7 444 146 BF320729
c 25 35.4 3.7 355 20 A108821
c 26 35.4 3.7 484 140 B6B61867
c 27 35.4 3.7 491 214 A0992818
c 28 35.4 3.7 544 144 BF113227
c 29 35.4 3.7 576 210 A0692130
c 30 35.4 3.7 609 167 BE460767
c 31 35.4 3.7 801 231 CNS04FNM
c 32 35.4 3.7 1101 229 CNS014JS
c 33 35.2 3.7 364 2 AA111946
c 34 35.2 3.7 711 232 PT006001R
c 35 35.2 3.7 874 150 BF576179
c 36 35.2 3.7 1201 229 CNS0160N
c 37 35 3.7 502 208 A0586301
c 38 35 3.7 925 229 CNS001BN
c 39 35 3.7 1101 229 CNS013VN
c 40 35 3.7 1201 229 CNS0108G
c 41 34.8 3.7 442 164 BE222338
c 42 34.8 3.7 879 231 CNS04PFG
c 43 34.8 3.7 916 230 CNS01W26
c 44 34.6 3.7 558 204 A0250056
c 45 34.6 3.7 713 148 BF422416

```

ALIGNMENTS

```

A1071988 UI-R-C2-n
A1100556 Drosophill
A1108811 Drosophill
AW463593 BP250013A
A0075795 CIT-HSP-2
BF320729 uz55907.Y
A108821 EST237112
B6B61867 UI-M-A00-
A0992818 RPCI-23-2
BF113227 EST40817
A0692130 HS_5328_B
BE460767 EST412186
A1288581 Tetradon
A1104242 Drosophill
AA111946 zn53e09.r
AL447600 Parametciu
BF576179 602132616
A1106145 Drosophill
A0586301 RPCI-11-4
A1074475 Drosophill
A1102077 Drosophill
A1098650 Drosophill
BE222338 hu10f11.x
A1288205 Tetradon
A1169863 Tetradon
A0250056 T3C5-Sp6
BF422416 Lr_ad_202

```

```

RESULT 1
LOCUS CNS005TE/c
DEFINITION Drosophila melanogaster genome survey sequence TET3 end of BAC #
BACR12K22 of RPCI-98 library from Drosophila melanogaster (fruit
fly), genomic survey sequence.
ACCESSION AL060767.1 GI:4943573
VERSION 1
KEYWORDS GSS.
SOURCE fruit fly.
ORGANISM Drosophila melanogaster
Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
Pterygota; Neoptera; Endopterygota; Diptera; Brachycera;
Muscomorpha; Ephydroidea; Drosophilidae; Drosophila.
REFERENCE 1 (bases 1 to 997)
AUTHORS Direct Submission
TITLE Submitted (02-JUN-1999) Genoscope - Centre National de Sequenage :
JOURNAL BP 191 91006 EVRY cedex - FRANCE (E-mail : seqref@genoscope.cns.fr
- Web : www.genoscope.cns.fr)
COMMENT Determination of this BAC-end sequence was carried out as part of a
collaboration with the Berkeley Drosophila Genome Project (BDGP).
The BDGP is constructing a physical map of the Drosophila
melanogaster genome using these BACs. For further information
please see http://www.fruitfly.org The BDGP Drosophila
melanogaster BAC library was prepared by Kazutoyo Osoegawa and
Aaron Mammose in Pieter de Jong's laboratory in the Department of
Cancer Genetics at the Roswell Park Cancer Institute in Buffalo,
NY. The library is named RPCI-98 and was constructed by partial
EcoRI digestion of Drosophila DNA provided by the BDGP from the
isogenic strain y2; cn bw sp. the same strain used for the BDGP's
p1 and Est libraries. A more detailed description of the library
and how to order individual BAC clones, the entire library, or
filters for hybridization from the BACPAC Resource Center can be
found at http://bacpac.med.buffalo.edu/drosophila_bac.htm.

```

FEATURES

```

source
1..997
/organism="Drosophila melanogaster"
/db_xref="taxon:7227"
/clone_lib="RPCI-98"
/clone="BACR12K22"

```

```

BASE COUNT 89 a 99 c 13 g 258 t 538 others
ORIGIN

```

Query Match

```

5.1% Score 48; DB 229; Length 997;
Best Local Similarity 18.8%; Pred. No. 0.0034;
Matches 72; Conservative 124; Mismatches 187; Indels 0; Gaps 0;

```

```

QY 302 aatataccaacatatacgaacgctgagcttggaaacaaagaatgctgtacattatg 361
:| | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 816 RARARCGRAAARARARARARARARARARARARARARARARARARARARARAR
362 atgcgaacaaagatgcgcgattattatggccacatgcacaaacaaagaagactaa 421
| | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 756 ARARAGARARARARARARARARARARARARARARARARARARARARARAGAA
422 ctaccatgataagctgagaaatgagatgagctctgcgcgtgtaagtaagccgg 481
: | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 696 RRRRGRGRRGAGARRGRRGRRGARRRRRRGRRRRRRRRRRRRRRRRRRRRRR
482 gggatttttctatgycgaacgctgactgctcatgcatgcatggaagaattctgctt 541
| | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 636 GARRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRAGAR
542 tggagacgcgcgaactcagacacacactacagattatgattatgacgcgaagaatg 601
| | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 576 RGRRRRRARARARARARARARARARARARARARARARARARARARARARARAR
602 cagaagcgaagctgcgcgagctcattcgtgaacaaagacattgaagtgcgcgt 661
: | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 516 RRRRRRGARRRRGRRRRRAGARRRRARARARARARARARARARARARARARAR
662 ctattccagaagcgcatacagt 684
| | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 456 NNNNNNNNNNNNNAGCMTNNNT 434

```

```

RESULT 2
LOCUS CNS016E2/c
DEFINITION Drosophila melanogaster genome survey sequence T7 end of BAC
BACN15A12 of DrosBAC library from Drosophila melanogaster (fruit
fly), genomic survey sequence.
ACCESSION AL106628.1 GI:5622852
VERSION 1
KEYWORDS GSS.
SOURCE fruit fly.
ORGANISM Drosophila melanogaster
Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
Pterygota; Neoptera; Endopterygota; Diptera; Brachycera;
Muscomorpha; Ephydroidea; Drosophilidae; Drosophila.
REFERENCE 1 (bases 1 to 1204)
AUTHORS Direct Submission
TITLE Submitted (23-JUL-1999) Genoscope - Centre National de Sequenage :
JOURNAL BP 191 91006 EVRY cedex - FRANCE (E-mail : seqref@genoscope.cns.fr
- Web : www.genoscope.cns.fr)
COMMENT Determination of this BAC-end sequence was carried out as part of a
collaboration with the European Drosophila Genome Project (EDGP) -
http://www.edgp.ebi.ac.uk -. This Drosophila melanogaster BAC
library (Dros BAC) was made by Alain Billand at CEPH (Centre
d'Etude du Polymorphisme Humain) with funding provided by a MRC
project grant. The DNA was prepared from embryos by Alain Bucheton
and Genevieve Payan. It has been constructed in the vector
pbeloBAC11.

```

FEATURES

```

source
1..1204
/organism="Drosophila melanogaster"
/plasmid="pbeloBAC11"
/db_xref="taxon:7227"
/clone_lib="DrosBAC"
/clone="BACN15A12"

```

```

/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="Plate2256 Col-5 Row=0"
/clone_1fp="CT Approved Human Genomic Sperm Library D
/sex="male"
/notes="Organ: sperm; Vector: pBeloBAC11; BAC Clones in

```

Query Match	4.18;	Score 38.6;	DB 232;	Length 1095;
Best Local Similarity	51.8%;	Pred. No. 2.1;		

[illegible]

RESULT	5
A0675159/c	
LOCUS	A0675159 419 bp DNA
DEFINITION	HS-2162_A1 G09_T7C Ctr Approved Human Genomic Sperm Library D
ACCESSION	sapiens genomic clone Plate=2162 COL17 Row=M, DNA sequence.
VERSION	A0675159
KEYWORDS	GSS.
SOURCE	human.
ORIGIN	

REFERENCE
AUTHORS
Mahairas,G.G., Wallace,J.C., Smith,K., Swartzell,S., Holzman,T.,
1 (bases 1 to 419)
Mannabild, Eucheria, Filinices, Cactaria, Monasterio, Mexico.

TITLE	Sequence-tagged connectors: A sequence approach to mapping and scanning the human genome
JOURNAL	Proc. Natl. Acad. Sci. U. S. A. 96 (17), 9739-9744 (1999)
MEDLINE	99380589
COMMENT	Contact: Mahairas GG, Wallace JC, Hood L

401 Queen Anne Avenue North, Seattle, WA 98109, USA
Tel: (206) 616-3618
Fax: (206) 616-3887
Email: jwallace@u.washington.edu
Clones may be purchased from Research Genetics (info@resgen.com)
BAC end Web server: <http://www.hsc.washington.edu>
Plate: 2162 row: M column: 17
Seq primer: T7
Class: BAC ends
High quality sequence stop: 419.

FEATURES	source	location/Qualifiers
1.	.419	
	/organism="Homo sapiens"	
	/db_xref="taxon:9606"	
	/clone="Plate2162 Col=17 Row=M"	
	/clone.lib="CIR Approved Human Genomic Sperm Library D"	
	/sex="male"	
	/note="organ: sperm; Vector: pBelBAC11; BAC Clones in E-Coli DH10B"	
BASE COUNT	110 a	99 c 77 g 133 t
ORIGIN		

[illegible]

QY	716	cgacatctgattggaatgagcgtctactcttlltcogryggsggaatggaactcaagctcagca	775
Db	239	GGACAGTGCACGAAATCCCAACGCTTTAGGGGGGATCAAGTATCATATTTGGCCATAATG	240
QY	776	gcttaagcagcaaaacacatcttcctctctatcagtgatgattgaagggagggccgtatga	835
Db	239	TATTTATGCTGCGCAAGGTATGGGGTGACAGTGAAGGGTGTAAGACTGGGGGCCAGGGGTGAGG	180
QY	836	tctctcgagatgattgtctatccctcttcaaaaaa	867
Db	179	TACCACTGATGTGTCAATATGACTTGCATATAATA	148

RESULT	6
AA550191/c	
LOCUS	AA550191 395 bp DNA
DEFINITION	1307m3 gmbpFHB3.1, G. Roman Reddy Plasmidium falciparum genomic clone 1307m, DNA sequence.
DATE	05-DEC-2000

ACCESSION	AA550191	GI:2320443
VERSION	AA550191.1	
KEYWORDS	GSS.	
SOURCE	Malaria parasite <i>P. falciparum</i> .	
ORGANISM	Plasmodium falciparum	
REFERENCE	Eukaryota; Alveolata; Apicomplexa; Haemosporida; Plasmodium	
	1 (bases 1 to 395)	

AUTHORS
Camejo, J.B., Arnott, D.E., Boulton, P., Ciardadadze, L., Christococci, A.,
Coppel, R., Comman, A., Craig, A., Fischer, K., Foster, J., Goodman, N.,
Hinderberg, K., Holder, A.A., Holt, D., Kemp, D., Lanzier, M., Lim, A.,
Newbold, C., Ravetch, J.V., Reddy, G.R., Rubio, J., Schuster, S.M., Su,
X.-Z., Thompson, J.K., Vital, F., Wellens, T.E. and Werner, E.
TITLE
Current status of the plasmodium falciparum genome project
JOURNAL
Mol. Biochem. Parasitol. 79, 1-12 (1996)

COMMENT

Contact: Dame JB
Dept. of Pathobiology, College of Veterinary Medicine
University of Florida
2015 SW 23rd Avenue, Bldg 1017, Gainesville, FL 32611, USA
Tel: 352 392 4700
Tel: 352 392 4700
Fax: 352 392 9704
Email: damej@mail.vetmed.ufl.edu
Seq primer: T3
Class: shotgun.

```

FEATURES
    source
        location/Qualifiers
            1..395
            /organism="Plasmodium falciparum"
            /db_xref="taxon:5833"
            /clone="1307m"
            /clone_id="gmbpFHB3.1, G. Roman Reddy"
            /lab_host="E. coli XL1-blue"
            /note="vector: pBluescript SK(+); Genomic DNA, from
            asynchronous blood stage parasites of the cloned Hondurana
            HB3 isolate cultured in vitro, was digested with mung bean
            nuclease in the presence of 30% formamide at 50°C (Verma
            , K.D., Imberiski, R.B., and McCutchan, T.F., 1988. Nucleic
            Acids Research 16:6883-6896). The ends of the fragments
            were ligated to EcoR V-cleaved and dephosphorylated
            pBluescript SK(+). Recombinant plasmids transformed E.
            coli XL1-Blue."

```

Query Match	4.0%	Score 38.2;	DB 201;	Length 395;
Best Local Similarity	49.5%;	Pred. No. 2.1;		
Matches 97;	Conservative 0;	Mismatches 99;	Indels 0;	Gaps 0.
Qy	257	tagatgctgcacgaactatctcttcagtgatgcgaatgaataatgccaacataac	316	
Db	389	TTAATATACATATAGTTCTTCTTATTAACAATAATTCAGATGCAACAGTTTTCACGTC	330	
Qy	317	atgaacacgctgagcttcggaacacagatgctcggtacatattatgtattgccaaaaagatg	376	

Oy	680	cagttacaccatgaacaacattgagaattgtgccttaccaacgacatttgatcgacttact	739
Db	255	AGCGAAGATGCAAGAAGAAAATTGCTGTATTAAACAAAGAAATCTTCAGAAAGCGTTTGG	314
Oy	740	tctcagtgggaataatggaacttatccgatcatcgaacttaagcagcctaagcagcaaaaaccttc	799
Db	315	AATTTCTCTTACGCACGAAGCATTTTTCMAATGCACAAGCAGGAAACTTGAAAAACAGTTTC	374
Oy	800	tctctacagtgtgattgaagg	822
Db	375	ATGATATGGCTTTGATGGTTGAG	397
RESULT	8		
LOCUS	AQI37339	375 bp	DNA GSS 24-SEP-1998
DEFINITION	HS.3058.B1.E03.MK.CIT Approved Human Genomic Sperm Library D Homo sapiens genomic clone Plate=3058 COL=5 Row=J, DNA sequence.		
ACCESSION	AQI37339		
VERSION	AQI37339.1	GI:3528185	
KEYWORDS	GSS.		
SOURCE	human.		
ORGANISM	Homo sapiens		
REFERENCE	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.		
AUTHORS	1 (bases 1 to 375) Mahalir,G., Wallace,J.C., Smith,K., Swartzell,S., Holzman,T., Keller,A., Shaker,R., Furlong,J., Young,J., Zhao,S., Adams,M.D. and Hood,L.		
TITLE	Sequence-tagged connectors: A sequence approach to mapping and scanning the human genome		
JOURNAL	Proc. Natl. Acad. Sci. U. S. A. 96 (17), 9739-9744 (1999)		
MEDLINE	99380589		
COMMENT	Contact: Mahaliras GG, Wallace JC, Hood L High Throughput Sequencing Center University of Washington 401 Queen Anne Avenue North, Seattle, WA 98109, USA Tel: (206) 616-3618 Fax: (206) 616-3887 Email: jwallace@u.washington.edu Sequence Tagged Connector Plate: 3058 row: J column: 5 Class: BAC ends High quality sequence stop: 375. Location/Qualifiers 1..375 /organism="Homo sapiens" /db_xref="taxon:9606" /clone_plate=3058 COL=5 ROW=J" /clone_lib="CIT Approved Human Genomic Sperm Library D" sex="male" /note="Organ: sperm; Vector: pBelBAC11; BAC Clones In E-Coli DH10B"		
CASE COUNT	121 a	71 c	67 g 115 t 1 others
ORIGIN			
Query Match	4.0%;	Score 38;	DB 202; Length 375;
Best Local Similarity	49.0%;	Pred. No. 2.3;	
Matches 101; Conservative	0;	Mismatches 105;	Indels 0; Gaps 0;
Oy	739	tttcgtggggaatggaacttatcagatcacgacgccaagcttaaagcagcaaacacattc	798
Db	55	TTTTGGTATNGTGTGTAATAATTATGAACAATCCCTTAATAAAGCAATTATAGATGCAAGTC	114
Oy	799	ctctctacagtgtgattgaaggagggccgatgatctctcgttgtgattgtctactcct	858
Db	115	ATCTGAAGTAGTACGCCCTTATGACAGCCTTTGAAACGATAGACATCTTCATGTATTATATG	174
Oy	859	tccaagaagaagatatattgtgtgtgcttaacgctcttgagaattaaacgcgaagga	918
Db	175	CTAATTAAGGAAAAAATCAGTTTCTTGACTGTGTATAGACGATCATCTCAATTAGCCAT	234

[illegible]

Db	373	GTGGCATCTAGCTTAACGAGGAGATTGAAGACAGACTCAATCATAGTGTGTATCAAAATGA	432
QY	591	ccgaaagatgacgaaggaagcagctgcgcgagcttcacatcgtgaaagaagacattgaatgt	650
Db	433	ACAAATAATGAAGAGTTGTGTGTCTTGCTTGCTGTCAGATGTAAGACAAAGTTCTACAGAC	492
QY	651	agaagtcctcgtatctccagaacggcataca	681
Db	493	AGCGACCTTAGTATTCAGACTGATTGCA	523
RESULT	10		
LOCUS	AZ209683		
DEFINITION	sp_0138.AL.A02_T7A Strongylocentrotus purpuratus, purple sea urchin	GSS	31-AUG-2000
ACCESSION	AZ209683		
VERSION	AZ209683		
KEYWORDS	clone plate=138 Col=3 Row=A, DNA sequence.		
SOURCE	Strongylocentrotus purpuratus.		
ORGANISM	Strongylocentrotus purpuratus.		
REFERENCE	Strongylocentrotus purpuratus; Eukaryota; Metazoa; Echinodermata; Eleutherozoa; Echinozoa; Echinoidea; Euechinoidea; Echinacea; Echinoidea; Strongylocentrotidae; Strongylocentrotus.		
AUTHORS	1 (bases 1 to 889) Cameron,R.A., Mahairas,G., Rast,J.P., Martinez,P., Biondi,T.R., Swartzell,S., Wallace,J.C., Poustka,A.J., Livingston,B.T., Wray,G.A., Ettensohn,C.A., Lehrach,H., Britten,R.J., Davidson,E.H. and Hood,L.		
TITLE	A sea urchin genome project: Sequence scan, virtual map, and additional resources		
JOURNAL	Proc. Natl. Acad. Sci. U. S. A. 97 (17), 9514-9518 (2000)		
MEDLINE	20402566		
COMMENT	Contact: Cameron, RA, Davidson, EH, Hood, L Division of Biology 156-29 California Institute of Technology Pasadena California 91125, USA Tel: (626) 395-8421 Fax: (626) 793-3047 Email: acameron@caltech.edu Plate: 138 row: A column: 3 Seq primer: T7 Class: BAC ends High quality sequence stop: 889.		
FEATURES	Location/Qualifiers		
source	1..889 /organism="Strongylocentrotus purpuratus" /db_xref="taxon:7668" /clone_plate=138 Col=3 Row=A" /clone_lib="Strongylocentrotus purpuratus, purple sea urchin, sperm genomic BAC library" /note="Organ: sperm; Vector: BACe3.6; BAC clones in E.Coli DH10B"		
BASE COUNT	267 a 188 c 134 g 300 t		
ORIGIN			
Query Match	4.0%;	Score 37.6;	DB 217; Length 889;
Best Local Similarity	50.6%;	Pred. No. 3.9;	
Matches	91; Conservative	0; Mismatches	89; Indels 0; Gaps 0;
QY	586	tatgaccggaagaatgcagaagcgaacgtgcgcgagcttcacatcgtgaaagaagacattgaa	645
Db	163	TACGTCCTTAATGCCACTTAACCCATACACGCGGAAAGACACTGACAGAGATGATTGA	222
QY	646	gtgatagaagtlccgctctattccagaacggcatacagttccaccatgaaagaattgagat	705
Db	223	ATCCAACTCGATCTGACTCTTGCCAGATTAACAAAGATTAATTAATTAATTAATTAACGTGA	282
QY	706	tgtccataacgacatgattgtaatgcgttcacttctcggtggggaatggaactatca	765
Db	283	TTCGACACAAAGAAATTCGATTATCACTAGTATTAATTAATGCTGTGATTTGGATTATCA	342

RESULT 11
 LOCUS A0842470 799 bp DNA GSS 12-MAY-2000
 DEFINITION CP61126A CP10WAGDNA1 Cryptosporidium parvum genomic, DNA sequence.
 ACCESSION A0842470
 VERSION A0842470.1 GI:6003998
 KEYWORDS GSS.
 SOURCE Cryptosporidium parvum.
 ORGANISM Eukaryota; Alveolata; Apicomplexa; Coccidia; Eimeriida;
 Cryptosporidiidae; Cryptosporidium.
 REFERENCE Strong, W.B. and Nelson, R.G.
 TITLE Preliminary profile of the Cryptosporidium parvum genome: an
 expressed sequence tag and genome survey sequence analysis
 JOURNAL Mcl. Biochem. Parasitol. 107 (1), 1-32 (2000)
 MEDLINE 20183851
 COMMENT Contact: Nelson, R. G.
 Depts. of Medicine & Pharmaceutical Chemistry
 San Francisco General Hospital-University of California, San
 Francisco
 Box 0811, San Francisco, CA 94143-0811, USA
 Tel: 415 206 8846
 Fax: 415 206 3353
 Email: malarielista.ucsf.edu
 For Annotation Data see <http://medsfgh.ucsf.edu/id/cpTags/home.html>
 Seq primer: M13(-21) Forward
 Class: shotgun.
 Location/Qualifiers
 1. 799
 /organism="Cryptosporidium parvum"
 /strain="IOWA"
 /db_xref="taxon:5807"
 /clone_11b="CP10WAGDNA1"
 /lab_host="E. coli XL2 Blue MRF"
 /note="Vector: pBluescript II (SK-); Site 1: ECoRV; C.
 parvum (IOWA isolate) genomic DNA was hydrodynamically
 sheared to produce fragments having a light size
 distribution between 2-4 kb by Dr. Yvonne Phorntenson of
 the Stanford DNA Sequencing and Technology Center
 (<http://sequence-www.stanford.edu/group/techdev/shear.htm>
). The randomly sheared gDNA was chromatographed on
 Sephacryl S-400 to remove any small fragments and DNA
 eluting in the void volume was subcloned into an EcoR
 V-digested, alkaline phosphatase-treated pBluescript II
 (SK-) vector and transformed into E. coli strain XL2 Blue
 MRF'. Recombinant clones from the first plating of the
 library were selected for sequence analysis using T3 and
 T7 primers."
 BASE COUNT 320 a 71 c 114 g 293 t 1 others
 ORIGIN
 Query Match 3.9%; Score 37; DB 212; Length 799;
 Best Local Similarity 51.5%; Pred. No. 5.7;
 Matches 85; Conservative 0; Mismatches 80; Indels 0; Gaps 0;
 Oy 252 aatatgagtcgacgaactatctgltcagtgatccgaatgtaataatgca 311
 ||||| || ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
 Db 397 AATATTTGGAATTTAATTCGATGTAACCTTTATTACTATTAATAAATCAAGATACAA 456
 ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
 Oy 312 catcagaataaagtgagcttgtaaaacgaatgctgacattatgttgccaa 371
 ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
 Db 457 AATTAACCAACAAATTTGAGGAGAAAGTACATTCAGGAATCTTAATGAATTAATAAA 516
 ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
 Oy 372 agatgcgagatattatgagccacaatgacaacaagaaga 416
 ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
 Db 517 AGATTGTGTAAGTATTTTGATTTGTAAGAAATTAATGAGAAA 561
 ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
 RESULT 12

CNS02H03/C
 LOCUS CNS02H03 1079 bp DNA GSS 13-MAY-2000
 DEFINITION Tetraodon nigroviridis genome survey sequence T7 end of clone
 138N19 of library G from Tetraodon nigroviridis, genomic survey
 sequence.
 ACCESSION AL197004.1 GI:7835154
 VERSION AL197004.1 GI:7835154
 KEYWORDS GSS; genome survey sequence.
 SOURCE Tetraodon nigroviridis.
 ORGANISM Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Actinopterygii; Neopterygii; Teleostei; Euteleostei; Neoteleostei;
 Eupercerygii; Ctenosquamata; Acanthomorpha; Eucanthomorpha;
 Holacanthopterygii; Acanthopterygii; Perciformes;
 Tetraodontiformes; Tetraodontidae; Tetraodontidae; Tetraodon.
 REFERENCE Roest-Crolius, H., Jaillon, O., Dasilva, C., Fizames, C., Fisher, C.,
 Bouneau, L., Billault, A., Quetier, F., Saurin, W., Bernot, A. and
 Weissenbach, J.
 TITLE Characterization and repeat analysis of the compact genome of the
 freshwater pufferfish Tetraodon nigroviridis
 JOURNAL Unpublished
 REFERENCE 2 (bases 1 to 1079)
 AUTHORS Roest-Crolius, H., Jaillon, O., Dasilva, C., Bouneau, L., Fisher, C.,
 Bernot, A., Fizames, C., Wincker, P., Brotier, P., Quetier, F.,
 Saurin, W. and Weissenbach, J.
 TITLE Human gene number estimate provided by genome wide analysis using
 Tetraodon nigroviridis DNA sequence
 JOURNAL Unpublished
 REFERENCE 3 (bases 1 to 1079)
 AUTHORS Genoscope.
 TITLE Direct Submission
 JOURNAL Submitted (12-APR-2000) to the EMBL/Genbank/DBJ databases
 COMMENT This sequence is a single read and was generated as part of a large
 scale clone-end sequencing project of the Tetraodon nigroviridis
 genome. For more information, please take a look at
<http://www.genoscope.cns.fr/Tetraodon>.
 Location/Qualifiers
 1. 1079
 /organism="Tetraodon nigroviridis"
 /db_xref="taxon:99883"
 /clone_11b="G"
 /note="Genoscope sequence ID : C0AG138G10LP1-end : T7"
 BASE COUNT 293 a 191 c 203 g 386 t 6 others
 ORIGIN
 Query Match 3.9%; Score 37; DB 230; Length 1079;
 Best Local Similarity 50.3%; Pred. No. 6.3;
 Matches 91; Conservative 0; Mismatches 90; Indels 0; Gaps 0;
 Oy 243 gcttacaanaatattagatgctgcagcaactatctgltcagtgatccgaatgatga 302
 ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
 Db 608 CGTGAATAAATGATCCAAACAGACGAGCAACAAATCTGATTCATTTGCTTTAATA 549
 ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
 Oy 303 atatgcacacataatgaataaagctgagcttggaataaagcaatgctgtacattatga 362
 ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
 Db 548 ATTAATTAATAAATAATTAATAATTAATAATTAATAATTAATAATTAATAATTAATA 489
 ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
 Oy 363 ttgcacaaagatgcgagatattatgacacaaatgacaacaagaagaactaac 422
 ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
 Db 488 ACTCTTAACGTATATGAGATTAACCATGTCATGATGATGATGATGATGATGATGATG 429
 ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
 Oy 423 t 423
 ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
 Db 428 t 428
 ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
 RESULT 13
 BF727776 411 bp mRNA EST 09-JAN-2001
 LOCUS BF727776
 DEFINITION 1000052F04.x2 1000 - Unigene 1 from Maize Genome Project Zea mays

Aaron Mammoser in Pieter de Jong's laboratory in the Department of Cancer Genetics at the Roswell Park Cancer Institute in Buffalo, NY. The library is named Rpci-98 and was constructed by partial EcoRI digestion of Drosophila DNA provided by the BDGP from the isogenic strain Y2; cn bw sp, the same strain used for the BDGP's P1 and EST libraries. A more detailed description of the library and how to order individual BAC clones, the entire library, or filters for hybridization from the BACPAC Resource Center can be found at http://bacpac.med.buffalo.edu/drosophila_bac.htm.

FEATURES

source

1. 1101

/organism="Drosophila melanogaster"

/db_xref="taxon:7227"

/clone_lib="Rpci-98"

/clone="BACR08K10"

/note="end : TET3"

BASE COUNT 201 a 64 c 131 g 202 t 503 others

ORIGIN

Query Match 3.9%; Score 36.6; DB 229; Length 1101;
Best Local Similarity 12.8%; Pred. No. 8.3; Mismatches 190; Indels 1; Gaps 1;
Matches 62; Conservative 230;

QY 342 agaagtcgtgtacatgatgccaagaagtcgagatattatgaccacatgc 401
DB 1089 RDTRKDDMDWTWKWTKWMDRDRWAGDADRWAMDGAGTWTATWMMWMTATWDTW 1030
QY 402 aacacacaaaggaactaactacatgataagcgtggaagaatggagctctgcg 461
DB 1029 WDKMWWMTAKTDTWTAMTWRADWAGRDGAGKRDADADGAGRDRGGRKRD 970
QY 462 ccgtgtaagtaagccgaggatcttctatgccaagcgtactgcatgcgac 521
DB 969 KKRKRGDDKKKGKKKAKAKATKWWMDWMDKMDKMDGAKRKRADDDGAGDKD 910
QY 522 tggaaagaagatctctgcttggaagcagacaactcagacacacacattata 581
DB 909 DGGKADDDDTGTDKDDKDKWMDKAKGTWGDATWMAATDMMWGMADADWMTWDA 850
QY 582 tgaattgaccgaagaagtcggaagcagcgtgcgactctcgtgaaagaagcat 641
DB 849 AADDMMADDKMDAMAKMD-AMAWGARTADRRWGDGAGKRGAKRRRKRADDKRDA 791
QY 642 tgaagtgaatagagtcctcctatccagaacgcatacagtcacacacaaatga 701
DB 790 ADDRDAATWTTWTTTTRDTDDMKKTDTWRMAADRTWDRDDDDDRAGTAGKMR 731
QY 702 ggaattgcttaacagacattgatgaatgccttacttctcgtyggggaatgaact 761
DB 730 TWRKRKRDRDTRWDADADDDRRRGDGDAGKGTGKRKRDRDRATWDRDADW 671
QY 762 atcagagtcagcagcttaagcagaacacacatcctctatcagtgatgaag 821
DB 670 WADAAWTTTDTDTDDWDRRRKARRRRRTTAAAMDWMTWAKMDWMTKTRADR 611
QY 822 gga 824
DB 610 WDR 608

Search completed: May 17, 2001, 15:08:04
Job time: 11490 sec

OM of: US-09-462-846-2 to: N_Geneseq_0401.* out_format : pfs

Date: May 17, 2001 4:13 PM

About: Results were produced by the GenCore software, version 4.5,
Copyright (c) 1993-2000 Compugen Ltd.

Command line parameters:

```
-MODEL=frim+p2n.model -DRV=xlp
-O=/cgn2_1/USPTO.spool/US09462846/runat_17052001_115429_970/app_query.fasta.1.376
-DB=N_Geneseq_0401 -QFMT=fastp -SUFFIX=p2n.rng -GAPOP=12.000
-GAPEXT=4.000 -MINMATCH=0.100 -LOOPEXT=0.000 -LOOPEXT=0.000
-OGAPOP=4.500 -OGAPEXT=0.050 -XGAPOP=10.000 -XGAPEXT=0.500
-FGAPOP=6.000 -FGAPEXT=7.000 -YGAPOP=10.000 -YGAPEXT=0.500
-DELOP=6.000 -DELEXT=7.000 -START=1 -MATRIX=blotsum62
-TRANS=human40.cdi -LIST=45 -DOCALLIGN=200 -THR_SCORE=pct
-THR_MAX=100 -THR_MIN=0 -ALIGN=15 -MODE=LOCAL -OUTFMT=pfs
-NORM=ext -MILEN=0 -MAXLEN=2000000000
-USR=US09462846.@cgn1_1.198 -NCPU=6 -ICPU=3 -LONGLOG -NO_XLPHY
-WAIT -THREADS=1
```

Search information block:

```
Query: US-09-462-846-2
Query length: 315
Database: N_Geneseq_0401.*
Database sequences: 678276
Database length: 291890651
Search time (sec): 126.050000
```

Score list:

Sequence	Strid	Orig	ZScore	EScore	len	! Documentation
/SID6/gcgdata/geneseq/geneseqn/NA1999.DAT:X25733 +				1692.00	3405.99	8.9e-182
/SID6/gcgdata/geneseq/geneseqn/NA1999.DAT:X25735 +				978.50	1959.57	3.3e-101
/SID6/gcgdata/geneseq/geneseqn/NA1999.DAT:X25734 +				958.50	1919.06	5.9e-99
/SID6/gcgdata/geneseq/geneseqn/NA1999.DAT:X13110 +				874.00	1719.53	7.7e-88
/SID6/gcgdata/geneseq/geneseqn/NA1999.DAT:X52181 +				801.00	1644.98	1.1e-83
/SID6/gcgdata/geneseq/geneseqn/NA1999.DAT:X74477 +				766.50	1382.78	4.4e-69
/SID6/gcgdata/geneseq/geneseqn/NA1999.DAT:X80389 +				685.00	1355.83	1.4e-67
/SID6/gcgdata/geneseq/geneseqn/NA1999.DAT:X74326 +				619.00	1215.71	8.9e-60
/SID6/gcgdata/geneseq/geneseqn/NA1999.DAT:X20249 +				166.00	259.81	1.6e-06
/SID6/gcgdata/geneseq/geneseqn/NA1999.DAT:X20248 +				166.00	257.26	2.8e-05
/SID6/gcgdata/geneseq/geneseqn/NA2000.DAT:X52111 +				145.50	230.52	5.1e-06
/SID6/gcgdata/geneseq/geneseqn/NA2001.DAT:X02175 +				145.50	243.64	1.2e-05
/SID6/gcgdata/geneseq/geneseqn/NA2001.DAT:X02174 +				145.50	242.45	1.4e-05
/SID6/gcgdata/geneseq/geneseqn/NA2000.DAT:X54213 +				145.50	242.32	1.5e-05
/SID6/gcgdata/geneseq/geneseqn/NA1999.DAT:X20087 +				145.00	249.53	5.8e-06
/SID6/gcgdata/geneseq/geneseqn/NA1999.DAT:X43322 +				115.00	203.04	0.0023
/SID6/gcgdata/geneseq/geneseqn/NA2000.DAT:X43143 +				109.00	194.53	0.0068
/SID6/gcgdata/geneseq/geneseqn/NA1999.DAT:X42063 +				101.50	98.86	1.4e+03
/SID6/gcgdata/geneseq/geneseqn/NA1999.DAT:X74580 +				96.50	163.41	0.3656
/SID6/gcgdata/geneseq/geneseqn/NA1999.DAT:X70431 +				96.50	164.15	0.3323
/SID6/gcgdata/geneseq/geneseqn/NA1999.DAT:X83879 +				93.50	151.50	1.68
/SID6/gcgdata/geneseq/geneseqn/NA1999.DAT:X83879 +				93.50	151.50	1.68
/SID6/gcgdata/geneseq/geneseqn/NA1999.DAT:X71208 +				93.00	158.12	0.7203
/SID6/gcgdata/geneseq/geneseqn/NA1998.DAT:X16515 +				92.50	149.28	2.24
/SID6/gcgdata/geneseq/geneseqn/NA1999.DAT:X83876 +				92.50	149.28	2.24
/SID6/gcgdata/geneseq/geneseqn/NA2000.DAT:X23235 +				92.50	148.18	2.58
/SID6/gcgdata/geneseq/geneseqn/NA1999.DAT:X74512 +				92.00	141.72	5.91
/SID6/gcgdata/geneseq/geneseqn/NA1999.DAT:X13047 +				89.50	144.56	4.10
/SID6/gcgdata/geneseq/geneseqn/NA2000.DAT:X09255 +				89.50	137.18	10.57
/SID6/gcgdata/geneseq/geneseqn/NA2000.DAT:X50334 +				88.00	139.15	8.21
/SID6/gcgdata/geneseq/geneseqn/NA2000.DAT:X47636 +				87.50	140.11	6.81
/SID6/gcgdata/geneseq/geneseqn/NA1992.DAT:X22989 +				87.50	141.63	6.81
/SID6/gcgdata/geneseq/geneseqn/NA1999.DAT:X77722 +				87.50	127.63	35.99
/SID6/gcgdata/geneseq/geneseqn/NA2000.DAT:X05585 +				87.00	151.58	1.67
/SID6/gcgdata/geneseq/geneseqn/NA2000.DAT:X33226 +				87.00	147.13	2.95
/SID6/gcgdata/geneseq/geneseqn/NA2000.DAT:X24744 +				87.00	145.71	3.54
/SID6/gcgdata/geneseq/geneseqn/NA1998.DAT:X26292 +				87.00	141.69	5.92
/SID6/gcgdata/geneseq/geneseqn/NA1998.DAT:X52337 +				86.50	133.93	16.04
/SID6/gcgdata/geneseq/geneseqn/NA1998.DAT:X52153 +				86.50	117.22	136.66

```
/SID6/gcgdata/geneseq/geneseqn/NA2000.DAT:X49283 + 86.00 147.91 2.67 1182
/SID6/gcgdata/geneseq/geneseqn/NA2000.DAT:X36468 + 86.00 146.36 3.25 1361
/SID6/gcgdata/geneseq/geneseqn/NA1999.DAT:X26612 + 86.00 142.54 5.31 1929
/SID6/gcgdata/geneseq/geneseqn/NA1999.DAT:X37760 + 86.00 131.89 20.83 5100
/SID6/gcgdata/geneseq/geneseqn/NA1999.DAT:X13117 + 86.00 111.94 269.23 31517
```

```
seq_name: /SID6/gcgdata/geneseq/geneseqn/NA1999.DAT:X25733
```

seq_documentation_block:

```
ID X25733 standard; DNA; 945 BP.
AC X25733;
XX
XX
XX 21-MAY-1999 (first entry)
DE B subtilis cysteine protease Cpl coding sequence.
XX
XX Cysteine protease; Gram-positive bacteria; mutation; deletion;
XX proteolysis; expression; heterologous protein; ss.
XX
XX Bacillus subtilis.
XX
XX WO904016-A2.
XX
XX 28-JAN-1999.
XX
XX 14-JUL-1998; 98WO-US14529.
XX
XX 15-JUL-1997; 97EP-0305227.
XX
XX (GENV) GENENCOR INT BV.
XX (GENV) GENENCOR INT INC.
XX
XX Estell DA:
XX WPI: 1999-132260/11.
XX P-PSDB: W99371.
XX
XX Gram-positive bacterium with mutated or deleted gene for cysteine
XX protease 1, 2 or 3 - used to express proteins with reduced
XX proteolytic degradation, e.g. proteins, growth factors or enzymes
XX
XX Disclosure: Fig 1A-B; 31pp; English.
```

This sequence represents the *Bacillus subtilis* cysteine protease Cpl coding sequence. The invention relates to new Gram-positive bacteria, especially *Bacillus*, containing mutations in, or deletion of, at least part of the gene encoding cysteine protease (Cpl1, Cpl2 (X25734) or Cpl3 (X25735)), so that CP-related proteolytic activity is lost. The bacteria are used for expressing heterologous proteins, particularly hormones, growth factors, cytokines or especially enzymes, e.g. protease, carboxylase, lipase, isomerases (tracemase, epimerase, tautomerase or mutase), transferase, kinase and phosphatase. Cpl-3 are used e.g. in soaps, dishwashing compositions, contact lens cleaners or laundry detergents, also for peptide hydrolysis, waste or textile treatment, to cleave fusion proteins and as animal feed additive.

Sequence 945 BP; 288 A; 180 C; 232 G; 245 T; 0 other;

alignment_scores:

```
Quality: 1692.00 Length: 315
Ratio: 5.371 Gaps: 0
Percent Similarity: 100.000 Percent Identity: 100.000
```

alignment_block:

```
US-09-462-846-2 x X25733 ..
Align seg 1/1 to: X25733 from: 1 to: 945
```

```
1 MetTrrHrGlupRleuPhePhelysProValPheylsGluArglleTr 17
|||||
1 ATGACGACTGAACCGTATTTCAGGCTGTTTCAAGAAAGAAATTTG 50
```

```

17 pglYgLYThrAlaLeuAlaaspPhegLYTYrThrIleProSerGlnArgT 34
   |||||
51 GGGGGGAGCCGCTTACGTGATTTGGCTATACATTCCTGCACACGAA 100
   |||||
34 hrGlyGluCysTrpAlaPheAlaAlaHISglnsngLYglnSerVal 50
   |||||
101 CAGGGAGTGTGGCTTTGGCCCGCATCAAAATGGTCAAGCGTGT 150
   |||||
51 GlnsngLYmetTYrLYsgLYPheThrLeuSerGlnuLeuPrpGlnHSH 67
   |||||
151 CAAGACGGAATGTAAGAGGCTCACGCAATTAAGGGAACCTCA 200
   |||||
67 sArHISLeuPhegLYglnLeuGluCysParGpheProLeuLeuThrl 84
   |||||
201 CAGACATTTATTCGACAGCTTGAAAGGGACCGTTCCCTGCTTACAA 250
   |||||
84 ySIleLeuAlaaspGlnaspGlnaspLeuSerValGlnValHISProasp 100
   |||||
251 AAATATTAGATGCTGACACGAGACTTATCTCAGTGTCATCCGAATGAT 300
   |||||
101 GluTYrAlaasnIleHISglnsngLYglnLeuGluCysTrpHISGly 117
   |||||
301 GAATATGCCAATACATCATGAAACGGTACCTTGAAACACAGAAATGCTG 350
   |||||
117 pTYrIleIleaspCysGlnLYsaspAlaGluIleIleTYrGlyHISasna 134
   |||||
351 GTACATTTATGATGCCAAAAAGATGCCGAGATTATTTATGGCCACAATG 400
   |||||
134 IArHrThrLYsgLYglnLeuThrThrmetIleGluArgGlyGluTYrPasp 150
   |||||
401 CAACAAAGAAAGAAAGAACTAACCTAGATAGAGCGTGAGAAATGGAT 450
   |||||
151 GluLeuLeuArgArgValLYsValLYsProGlyaspPhePheTYrValPr 167
   |||||
451 GAGCTTGGCGCGGTAAAGTAAGCCGGGAGATTTTTCATATGTGCC 500
   |||||
167 oSerGlyThrValHISAlaIleGlyLYsgLYIleLeuAlaLeuGluThrg 184
   |||||
501 AAGGGGACTGTTCAATCGATGGAAGAAATTCCTTGTGAGACGC 550
   |||||
184 InGlnAsnSerAspThrThrTYrArgLeuTYrAspTYrAspArgLYsasp 200
   |||||
551 AGCAGAACTCAGACACACTACAGATTAATATTAATATGACCAAAAGAT 600
   |||||
201 AlAGluGlyLYsLeuArgGluLeuHISLeuLYsSerIleGluValIle 217
   |||||
601 GCAGAAAGCAAGCTGGCGAGCTTCACTGAAAGAACAGCATTGAACTGAT 650
   |||||
217 eGluValProSerIleProGluArgHISThrValHISHSglnGlnIleG 234
   |||||
651 AGAGGTCCTCGTATTCACAGACGCAATACAGTTCAACATGAACAAATG 700
   |||||
234 IuAspLeuLeuThrThrThrLeuIleGluCysAlaTYrPheSerValGly 250
   |||||
701 AGGATTTGCTTACAGACACTGATTAAGTGCCTTACTTTCGGGGG 750
   |||||
251 LYsTrpAsnLeuSerGlySerAlaSerLeuLYsGlnHISLYsProPheLe 267
   |||||
751 AAATGGAACCTTATCAGATCAGCAACCTTAAGCAACAAACCAATTCCT 800
   |||||
267 uLeuIleSerValIleGluGlyGluLYrGmetIleSerGlyGluTYrV 284
   |||||
801 TCTTATCAGTGTATGAAGGGAGGGCCGTATGATCTGTGGAGTATG 850
   |||||
284 aLYrProPheLYsLYsGlyAspHISmetLeuLeuProTYrGlyLeuGly 300
   |||||
851 TCTATCCTTTCAAAAAAGAGATCATATGTGCTGCTTACGCTTGCA 900
   |||||
301 GluPheLYsLeuGluGlyTYrAlaGluCysIleValSerHISLeu 315
   |||||
901 GAATTTAACTCGAAAGATATGACAGATGTATGCTCCATCTG 945

```

```

seq_name: /SIDS6/gcdata/geneseq/geneseqn/NA1999.DAT:X25735
seq_documentation_block:
ID: X25735 standard; DNA; 948 BP.
XX
XX
AC: X25735;
XX
XX
DT: 21-MAY-1999 (first entry)
XX
DE: B. subtilis cysteine protease CP3 coding sequence.
XX
KW: Cysteine protease; Gram-positive bacteria; mutation; deletion;
   proteolysis; expression; heterologous protein; ss.
XX
OS: Bacillus subtilis.
XX
PN: W09904016-A2.
XX
PD: 28-JAN-1999.
XX
PF: 14-JUL-1998; 98MO-US14529.
XX
PR: 15-JUL-1997; 97EP-0305227.
XX
PA: (GEMV ) GENENCOR INT BV.
   (GEMV ) GENENCOR INT INC.
XX
PI: Estell DA.
XX
DR: WPI: 1999-132260/11.
DR: P-PSDB: W99373.
XX
PT: Gram-positive bacterium with mutated or deleted gene for cysteine
   protease 1, 2 or 3 - used to express proteins with reduced
PT: proteolytic degradation, e.g. proteins, growth factors or enzymes
XX
PS: Disclosure; Fig 6A-B; 31pp; English.
XX
CC: This sequence represents the Bacillus subtilis cysteine protease CP3
   coding sequence. The invention relates to new gram-positive bacteria,
   especially Bacillus, containing mutations in, or deletion of, at least
   part of the gene encoding cysteine protease (CP1 (X25733), CP2 (X25734)
   or CP3), so that CP-related proteolytic activity is lost. The bacteria
   are used for expressing heterologous proteins, particularly hormones,
   growth factors, cytokines or especially enzymes, e.g. protease,
   carbohydrase, lipase, isomerases (racemase, epimerase, tautomerase or
   mutase), transferase, kinase and phosphatase. CP1-3 are used e.g. in
   CC: soaps, dishwashing compositions, contact lens cleaners or laundry
   CC: detergents, also for peptide hydrolysis, waste or textile treatment,
   CC: to cleave fusion proteins and as animal feed additive.
XX
SQ: Sequence 948 BP; 286 A; 192 C; 244 G; 226 T; 0 other;
align_scores:
  Quality: 978.50      Length: 316
  Ratio: 3.822         Gaps: 1
  Percent Similarity: 81.013   Percent Identity: 56.646
alignment_block:
US-09-462-846-2 x X25735 ..
Align seg 1/1 to: X25735 from: 1 to: 948
1 MetThrGluProLeuPhePheLYsProValPheLYsgluArgIleTr 17
   |||||
1 ATGACGCAATCACCGATTTTCTAACCCCTGTGTTTAAAGAAATCTG 50
   |||||
17 pglYgLYThrAlaLeuAlaasp...PhegLYTYrThrIleProSerGlna 33
   |||||
51 GGGGGGAGCCGCTTACGATAGATTTGGATACAGATTAATTCCTTCAAAAT 100
   |||||
33 rGThrGlyGluCysTrpAlaPheAlaAlaHISglnsngLYglnSerVal 49

```

101 CAACGGGGAATGCTGGGCCATTTCGCGTATTCACAAAGGACCGACACT 150
50 ValGlnasnGlyMetYrlySGlyPheThrLeuSerGlyLeuTrpGlnH 66
151 GTTCCAAATGAGCCCGTATAAAGGAAGACATATCATGAGCTTTGGGAGA 200
66 SHIARHISLeuPheGlyGlnGluGlyYAspArgPheProLeuLeuT 83
201 GCACCGTGAAGATTTCCGGCCGCTGAGAGGGGATCGTTTTCCGCTTCTGA 250
83 hrLySLeuAspAlaAspGlnAspLeuSerValGlnValHisProasn 99
251 CAAAGCTGCTGGATGTGAAGAAGATACGTCATTAAGTTCAACCTGAT 300
100 AspGluTyrAlaAsnIleHisGlnAsnGlyGluLeuGlyYsrThrGly 116
301 GATTACTATGCGGAGAAACGAAGAGGGAACCTCGCAACGGAAGAT 350
116 STRPTyrIleLeuAspGlyGlnYAspAlaGluIleIleTyrGlyHis 133
351 CTGGTACTATTTCACCTGTAAAGAAACGCAAGATCATTTAGGGCAT 400
133 snAlaThrThyLysGlyGluLeuThrThrMetIleGluArgGlyLys 149
401 CGGCCCGCTCAAAACCGAATCTGTCAATGATCAACACCGCTGACTGG 450
150 AspGluLeuLeuArgValYValYValYProGlyAspPheThrYrva 166
451 GAGGGCCGTGGCGAAGATCAAAATTAACCCGGGTATTTCTATTATGT 500
166 IProSerGlyThrValHisAlaIleGlyLysGlyIleLeuAlaLeuGly 183
501 GCGGAGCGGAACGCTGCACGCATGTGTCAAGAGGGCCCTGTTTAGA 550
183 hrGlnGlnAsnSerAspThrThrYrArgGluLeuYrAspYrAspArgLys 199
551 CTCAGCAAAATTCAGATGCCACATACCGGGGTAGCATTAATGACCGCTT 600
200 AspAlaGlnGlyLysLeuArgGlnLeuHisLeuLysLysSerIleGluVal 216
601 GATGCAACGGAAGATCGAGACCTTATTTGCCAAAGCGGTCAATGC 650
216 ILeuGluValProSerIleProGluArgHisThrValHisIleGluGln 233
651 CGCCGCGGTCCCCATGGGAGCGGATATATGATGAATCGACAAACAA 700
233 IeGluAspLeuLeuThrThrThrLeuIleGluLysAlaTyrPheSerVal 249
701 GAAAGGAATATCACTTAAACATTTGTGCCAAGGGAATATTTTCGCT 750
250 GlyLysTrpAsnLeuSerGlySerAlaSerLeuLysGlnGlnLysProph 266
751 TATAATATGGCAATCAATAGCGAAGCTGAATATGGCTCAGATGATCACT 800
266 eleuLeuLeuSerValIleGluGlyGlyArgMetLeuSerGlyGln 283
801 TGTGATTTGCCAGCTGATAGAGGAAGCGGTTTGTCTCAAGTATAGACA 850
283 YrValTyrProPheLysLysGlyAspHisMetLeuLeuProTyrGlyLeu 299
851 AAACATGTCGCGTCAAAAGAGGTATCTATTATTTGCCGCTCAATAG 900
300 GlyGluPheLysLeuGlyGlyTyrAlaGluLysIleValSerHisLeu 315
901 CCGGATTTTACATATAAAGACATTGTACCCCTTATCGGTCTCATAT 948
seq_name: /SIDS6/gcgcdata/geneseq/geneseqn/NA1999.DAT.X25734
seq_documentation_block:
ID X25734 standard; DNA; 945 BP.
XX
AC X25734;

```

XX 21-MAY-1999 (first entry)
DE
XX
XX B.subtilis cysteine protease CP2 coding sequence.
DE
XX
XX Cysteine protease; Gram-positive bacteria; mutation; deletion;
KW proteolysis; expression; heterologous protein; ss.
XX
XX Bacillus subtilis.
OS
XX
XX WO9904016-A2.
PM
XX
XX 28-JAN-1999.
PD
XX
XX 14-JUL-1998; 98WO-US14529.
XX
XX 15-JUL-1997; 97EP-0305227.
PR
XX
XX (GENM ) GENENCOR INT BV.
PA (GENM ) GENENCOR INT INC.
XX
XX
XX Estell DA.
PI
XX
XX WPI: 1999-132260/11.
DR
XX
XX P-PSDB: W99372.
XX
XX Gram-positive bacterium with mutated or deleted gene for cysteine
PT protease 1, 2 or 3 - used to express proteins with reduced
PT proteolytic degradation, e.g. proteins, growth factors or enzymes
XX
XX
XX Disclosure: Fig 5A-B; 31pp: English.
XX
XX
XX This sequence represents the Bacillus subtilis cysteine protease CP2
XX coding sequence. The invention relates to new Gram-positive bacteria,
XX especially Bacillus, containing mutations in, or deletion of, at least
XX part of the gene encoding cysteine protease (CP1) (X25733), CP2 or CP3
XX (X25733), so that CP-related proteolytic activity is lost. The bacteria
XX are used for expressing heterologous proteins, particularly hormones,
XX growth factors, cytokines or especially enzymes, e.g. protease,
XX carboxylidase, lipase, isomerases (racemase), epimerase, lactomerase or
XX mutase), transferase, kinase and phosphatase. CP1-3 are used e.g. in
XX soaps, dishwashing compositions, contact lens cleaners or laundry
XX detergents, also for peptide hydrolysis, waste or textile treatment,
XX to cleave fusion proteins and as animal feed additive.
XX
XX
XX Sequence 945 BP; 290 A; 189 C; 223 G; 243 T; 0 other;
SX

```

```

68 gHISLeuPheGlyLeuGlyLeuGlyAspArgPheProLeuThrLysI 85
204 AGAGATATTCGGGCTTCGGATGTGAAGTCTTCCTGCTGTGAAGC 253
85 lELeuAspAlaAspGlnAspLeuSerValGlnValHisProAsnAspGlu 101
254 TGTGGAGCCCAATATGATGCTCTCCGTAAGTCATCCCTGATGATGAT 303
102 TyrIleAsnIleHisGlnAsnGlyGlnLeuGlyLysThrGluCysTrpTy 118
304 TATGCAAAACGTCACGAAATGCGACCTGTGTAACCGAGTCTGTGTA 353
118 rIleIleAspCysGlnLysAspAlaGluIleIleTyrGlyHisAsnAla 135
354 TATCATTTGATTCGAAGATGACCGCAACTAATTTTGGACATCATTCGA 403
135 hrThrLysGlnGluLeuThrThrMetIleGluArgGlyGluTrpAspGlu 151
404 GCACAAAGGAAGAGTTCAAACACACGAATAGAAACGGGTGATGGACGG 453
152 LeuLeuArgArgValLysValLysProGlyAspPhePheTyrValProse 168
454 CTGCTGAGCCCAATCAATCAATCAAGCCGAGATTTCTTTATGTGCGAAG 503
168 rGlyThrValHisAlaIleGlyLysGlyIleLeuAlaLeuGlnThrGln 185
504 CGGTACACTCCTCATTTATGTAGGACCCCTGCTCTGAAATCAGC 553
185 lAsnAspArgPheThrThrArgLeuTyrAspTyrAspArgLysAspAla 201
554 AAAACCTGTATACACATATCGCTATACCATTTATGACCCCTGTAAGAC 603
202 GluGlyLysLeuArgGluLeuHisLeuLysSerIleGluValIleGlu 218
604 CAGGCGCAAAAGAACTCTTCATATAGAAAAGCCATGGAATCATCAAC 653
218 uValProSerIleProGluArgHisThrValHisHisGlnGlnIleGlu 235
654 GATACCCCATATCGATTAAGTCATACACCGGAAGCAAGAGTGGCTA 703
235 sPLeuLeuThrThrThrLeuIleGluCysAlaTyrPheSerValGlyLys 251
704 ACGGTGAGATCATTTGTTATGTGCAATCAATATTTCTGAGTGTACAA 753
252 TrpAsnLeuSerGlySerAlaSerLeuLysGlnGlnLysProPheLeu 268
754 TGGAAATATAGCGCGCCGCTGCTTCCTTCATATCAACCTATTTGCT 803
268 uIleSerValIleGlnGlyGlnGlyArgMetIleSerGlyGluTyrVal 285
804 GGGGAGTGTCTGAGCGGATCAGACGCAATCAATAAATGAATTCAGT 853
285 yTrpPheLysLysGlyAspHisMetLeuLeuProTyrGlyLeuGlyGln 301
854 ATGAATGCATGACAGGCTCACACTTATTTGCTGCGCATTTTGGAGAA 903
302 PheLysLeuGlnGlyTyrAlaGluCysIleValSerHis 314
904 TTTACATATGAAAGAACATGTGAATTCATGATATCTCAT 942
seq_name: /SID6/gcgdata/geneseq/geneseq/NAL199.DAT:XL13110
seq_documentation_block:
ID XL13110 standard; DNA: 12438 BP.
XX
AC XL13110;
XX
DT 19-MAR-1999 (first entry)
XX
DE Enterococcus faecalis genome contig SEQ ID NO:173.
XX
KW Enterococcus faecalis; contig; detection; Enterococcal infection;

```

```

KW vaccine; attenuation; computer readable medium; ds.
OS Enterococcus faecalis.
XX
PN MO9850555-A2.
XX
PD 12-NOV-1998.
XX
PF 04-MAY-1998; 98WO-US08985.
XX
PR 14-NOV-1997; 97US-0066009.
PR 06-MAY-1997; 97US-0044031.
PR 16-MAY-1997; 97US-0046555.
XX
PA (HUMA-) HUMAN GENOME SCI INC.
XX
PI Barash SC, Dillon PJ, Kunsch CA;
XX
DR WPI; 1999-045171/04.
XX
PT New Isolated Enterococcus faecalis polynucleotides and polypeptides
PT - used to develop products for the detection of Enterococcus and for
PT use in vaccines for prevention or attenuation of Enterococcus
PT infection.
XX
PS Claim 1; Page 963-969; 2084pp; English.
XX
CC A computer readable medium has been developed which has recorded on it
CC 982 nucleotide sequences isolated from the Enterococcus faecalis genome.
CC X12938 to X13191 represent these nucleotide sequences which are primary
CC nucleotide sequences, also known as contigs. The computer-based system
CC can identify fragments of the Enterococcus faecalis genome with
CC commercial importance. The products can be used to detect the presence
CC of Enterococcus faecalis in samples. They can also be used for
CC diagnosing Enterococcal infection in an animal and monitoring
CC progression of disease, and for identifying agents which can be used to
CC modulate the growth or pathogenicity of Enterococcus faecalis, or
CC another related organism, in vivo or in vitro. In particular the
CC polypeptides encoded by the Enterococcus faecalis nucleotide sequences
CC can be used in vaccines to prevent or attenuate an Enterococcal
CC infection.
XX
SO Sequence 12438 BP; 4071 A; 2218 C; 2424 G; 3708 T; 17 other;

alignment_scores:
Quality: 874.00 Length: 317
Ratio: 3.553 Gaps: 3
Percent Similarity: 77.603 Percent Identity: 53.628

alignment_block:
US-09-462-846-2 x XL13110 ..

Align seg 1/1. to: XL13110 from: 1 to: 12438

4 GIUPrOLeuphePheLysProValPheLysGluArgIleTrpGlyGlyTh 20
|||||
6276 GAACCATATTTTTCACACCTGTTTTCAGAAAATTTGGCGCGGAA 6325
20 rAlaLeuAlaAsp...PheGlyTyrThrIleProSerGlnArgThrGly 36
|||
6326 TCGTTCACACACCGCTATTGGTTTCGATTTCAGCGGATTAATTTGCTG 6375
36 lUcYTrpAlaPheAlaHisGlnAsnGlyGlnSerValIleGlnAsn 52
|||
6376 AAGATGGCAATCAGTCACATCAGATGCGCTTAGTACTGTTTAAAT 6425
53 GlyMetTyrLysGlyPheThrLeuSerGluLeuTrpGlnHisArgHis 69
|||
6426 GCGCAATTTAAAGGAAAAAATTTAGATGAATTTGGCGACACATCAAGA 6475
69 sLeuPheGlyGlnLeuGlyAspArgPheProLeuLeuThrLysIle 86
|||||

```

```

6476 GTTATTCGGCCATGCGAGGTGGCGTCTGCTTTCATTATTAAAGAAATTT 6525
86 euAspAlaAspGlnAspLeuSerValGlnValHisProAsnAspGluTyr 102
|||||.....|||||.....|||||.....
6526 TAGATGCTGAAGATGATTTATCCGTGCAAGTTCATCTGATGATCATAT 6575
103 AlaAsnIleHisGlnAsnGluValLeuGluTyrThrGluCysTrpTyrIle 119
.....|||||.....|||||.....|||||.....
6576 GGCCAAAGACATGAA...GGCGAGCTAGGCAAAAGTGAATGTTGGTACAT 6622
119 eileAspCysGlnLysAspAlaGluIleTyrGlnHisAsnIleThrT 136
|||||.....|||||.....|||||.....
6623 TATGATGAGAAAGCTGGCGCTGAATCATTTATGACATCAGCGAAAA 6672
136 hrlYsGluGluLeuThrThrMetIleGluArgGlyGluTrpAspGluLeu 152
|||||.....|||||.....|||||.....
6673 CAAAGAAAGAACTGCGAGAAAGATGAAAGAGTGGTGGGACAACTT 6722
153 leuArgArgValLysValLysProGlnAspPhePheTyrValProSerGlu 169
|||||.....|||||.....|||||.....
6723 TTAAAGAAAGTACAGTGAAGAAAGCGATTTTTATGATGACCAAGCGG 6772
169 yhrValHisAlaIleGlyLysGlyIleLeuAlaLeuGluThrGlnGln 186
|||||.....|||||.....|||||.....
6773 CACTATTCATGCGCATGGCTCGGTATTTTATTAGAAACGCAACAA 6822
186 snSerAspThrThrThrTyrArgLeuTyrAspTyrAspArgLysAspAlaGlu 202
|||||.....|||||.....|||||.....
6823 GCAGTCAATCTACGTATCGGTTTATGATTAATGATGACCAATGATCAA 6872
203 GlyLysLeuArgGluLeuHisLysSerIleGluValIleGluVal 219
|||||.....|||||.....|||||.....
6873 GGAAGAACCCGGGAATGTCATATTCACAAATCCATGATGATCAACCT 6922
219 lProSerIleProGluArgHisThrValHisIleGluGlnIleGluAspL 236
|||||.....|||||.....|||||.....
6923 TCCGCGAGAGATCCGACCTCTATTCACACAGAAATCAAGAACAT 6972
236 euleuThrThrThrIleuIleGluCysAlaTyrPheSerValGlyTyr 252
|||||.....|||||.....|||||.....
6973 CAAGCATTTCTTATGTAATAAAACAGATTTCTTATATCTATGAGTGG 7022
253 AsnLeuSerGlySerAlaSerLeuLysGlnGlnLysProPheLeuLeuI 269
|||||.....|||||.....|||||.....
7023 CGGTAGAGGGATTTCTAAAGTCAAAACAGACCTTATACATTAC 7072
269 eSerValIleGluGlyGluGlyArgMetIleSerGlyGlu..... 282
|||||.....|||||.....|||||.....
7073 AACAGTATTTGAAGTGTGGTCTGATTGATTACAGAAAGATCTGCAAAAG 7122
283 .....TyrValTyrProPheLysGlyAspHisMetLeuPro 296
|||||.....|||||.....|||||.....
7123 CAGATGTCGCTACTTTGATTTGAAAAAGGCGATAGCTTTATCTTCGG 7172
297 TyrGlyLeuGlyGluPheLysLeuGluGlyTyrAlaGluCysIleValSe 313
|||||.....|||||.....|||||.....
7173 ACAGATATTCGAGCTGGCGTTTGAAGTGAATCAAGCATATATAGCTTC 7222
313 r 313
7223 A 7223
seq_name: /SID56/gcdata/geneseq/geneseqn/NA1998.DAT:V52181
seq_documentation_block:
ID V52181 standard; DNA: 25002 BP.
XX V52181;
XX 23-OCT-1998 (first entry)
XX Streptococcus pneumoniae genome fragment SEQ ID NO:48.

```

```

KW Streptococcus pneumoniae; S. pneumoniae; genome; diagnosis; assay;
KM computer readable medium; vaccine; pharmaceutical composition; ds.
XX Streptococcus pneumoniae.
XX WC09818931-A2.
XX 07-MAY-1998.
XX 30-OCT-1997; 97WO-US19588.
XX 31-OCT-1996; 96US-0029960.
XX (HUMA-) HUMAN GENOME SCI INC.
XX Barash SC, Choi GH, Dillon PJ, Dougherty BA, Pannon M,
XX Kunsch CA, Rosen CA;
XX MPI; 1998-272225/24.
XX
XX Computer-readable medium with recorded Streptococcus pneumoniae
XX polynucleotide sequences - useful in diagnostic kits and assays, and
XX pharmaceutical compositions and vaccines for Streptococcus
XX pneumoniae
XX
XX Claim 1; Page 431-445; 1409pp; English.
XX
XX The present invention describes a computer readable medium which has
XX the nucleotide sequences SEQ ID NO:1 to 391 (V52134 to V52524) recorded
XX on it, or a representative fragment or a sequence at least 95% identical
XX to SEQ ID NO:1 to 391. The nucleotide sequences depicted in SEQ ID NO:1
XX to 391 (V52134 to V52524) are genomic fragments from Streptococcus
XX pneumoniae. The present invention also describes an isolated nucleic acid
XX molecule encoding a homologue of any of the fragments of the S.pneumoniae
XX genome (SEQ ID NO:1 to 391) where the nucleic acid molecule is produced
XX by a process comprising: (a) screening a genomic DNA library using as a
XX probe a target sequence defined by any of the sequences in SEQ ID NO:1
XX to 391, identifying members of the library which contain sequences
XX that hybridize to the target sequence and isolating the nucleic acid
XX molecules from the members; or (b) isolating mRNA, DNA or cDNA produced
XX from an organism, amplifying nucleic acid molecules whose nucleotide
XX sequence is homologous to amplification primers derived from the
XX CC fragment of the S. pneumoniae genome to prime the amplification and
XX CC isolating the amplified sequences. The computer readable medium can be
XX CC used in a computer-based system for identifying fragments of the
XX CC S. pneumoniae genome of commercial importance, or expressions of
XX CC fragments of the S. pneumoniae genome. Products from the present
XX CC invention can be used in diagnosis kits and assays, and pharmaceutical
XX CC compositions and vaccines for S. pneumoniae.
XX
XX Sequence 25002 BP; 7422 A; 5498 C; 4618 G; 7463 T; 1 other;
XX
XX
XX alignment_scores:
XX Quality: 841.00 Length: 316
XX Ratio: 3.461 Gaps: 5
XX Percent Similarity: 76.899 Percent Identity: 53.797
XX
XX alignment_block:
XX US-09-462-846-2 x V52181/rev ..
XX
XX Align seg 1/1 to reverse of: V52181 from: 1 to: 25002
XX
XX 3 ThrGluProLeuPhePheLysProValPheLysGluArgIleTrpGly 19
XX TCAGAACCATTTATTTTACAAATCAGATTATGCAAGAAAAAATCTGGGTGG 2332
XX
XX 19 yThrAlaLeuAlaAsp...PheGlyTyrThrIleProSerGlnArgThr 35
XX |::: ||| ||| ||||| ||||| ||||| ||||| |
XX |::: ||| ||| ||||| ||||| ||||| ||||| |
XX 2331 AGCCAAAGCTACGTATGATGTTGGCTAGCAGATCCCAAGTAAAAAATCG 2282
XX
XX 35 lYgluCysTrpAlaPheAlaHisGlnAsnGlyGlnSerValValGln 51
XX ||||| |||||.....||||| ||||| ||| |||

```

```

2281 GAGATATTGGCCATCTCAGCCCATCCAAATGGAGTCTTAAAGTGGC 2232
XX
52 AsnGlyMetTyrGlyPheThrLeuSerGluLeuTrpGluHisHisAr 68
XX
2231 AATGCTCTTACGAGGGAACAGATCTTGTCTACTTGTATGGGGAACACC 2182
XX
68 GHisLeuPheGlyGluLeuGluGlyAspArgPheProLeuLeuThrLysI 85
XX
2181 TGAATTATTGGCAATCTCCAGAACCTGTATTTCACATTTCACCAAGA 2132
XX
85 LeuAspAlaAspGlnAspLeuSerValGlnValHisProAsnAspGlu 101
XX
2131 TCCTGATGCCACAGCAGCTGCTCAGTGTCCAGTTGCCACCCAGACGATGCT 2082
XX
102 TyrAlaAsnIleHisGluAsnGlyLeuGluGlyLysThrGluCysTrpTyr 118
XX
2081 TATGAGACTCGAGCATGAA...GGCGAACTCGGAAACAGAAATGCTGTGA 2035
XX
118 rIleLeuAspCysGlnLysAspAlaGluIleIleTyrGlyHisAsnAlaT 135
XX
2034 CATATCGCAGCGGATGAAAGGTTCAGAGATTATCTACGGTCACAAATGCCA 1985
XX
135 hTrpLysGluGluLeuThrThrMetIleGluAspGlyGluTrpAspGlu 151
XX
1984 AGTCAAAAGAAAGAACTCCGCCAGCAATCCAGGACAGAACTGGGATGAC 1935
XX
152 LeuLeuArgArgValLysValLysProGlyAspPhePheTyrValProse 168
XX
1934 TTGTTGACCAAAAGTCCCTGTAAAGCTGAGAGATTCTTCTATGTACCAAG 1885
XX
168 rGlyThrValHisAlaIleGlyLysGlyIleLeuAlaLeuGluThrGln 185
XX
1884 CGGAGCTATGACGCTATCCGGGGGTATCTTAATCTTGAAACCAGC 1835
XX
185 InAsnSerAspThrThrTyrArgLeuTyrAspTyrAspArgLysAspAla 201
XX
1834 AGTCTAGCAGATCCACCACTATGCGCTGTATGACTTTCACCCAGAGATGAC 1785
XX
202 GluGlyLysLeuArgGluLeuHisLeuLysSerIleGluValIleG1 218
XX
1784 AAAGCAAACTTGCCTGAACTCATCTTGAATAATCCATGCTGTGA 1735
XX
218 uVal.....ProSerIleProGluArgHisThrValHisIleGluGln 233
XX
1734 CATGTGTGAGCCAGCAAAATAGCCGCTGTAACTGTT.....AAAG 1694
XX
233 IeGluAspLeuLeuThrThrThrLeuIleGluCysAlaTyrPheSerVal 249
XX
1693 CAGATGATTTGCGTTCACCTCTCTGTATCTATGATGATTTCTTCGCAGT 1644
XX
250 GlyLysTrpAsnLeuSerGlySerAlaSerLeuLysGlnGlnLysProp 266
XX
1643 TACAAGTGGGAAATTTACTGGAAAGTTGACTTTGAAAGAACACACTGACTA 1594
XX
266 eLeuLeuIleSerValIleGluGluGluArgMetIleSerGlyLur 283
XX
1593 CAGCTTATGAGTCTTGGCTGCTCAAGGTCAGCTAAGCTAGCTTGACGGGA 1544
XX
283 yValTyrProPheLysGlyAspHisMetLeuLeuProTyrGlyLeu 299
XX
1543 AAATATATCCAAATCAAAAGGACGACCTTATCTACCAAGATGATGT 1494
XX
300 GlyLuuLeuLysLeuGluGlyTyrAla..GluCysIleValSerHis 314
XX
1493 GAACTTGGAGCTCTGGAAGGGAAGGTTTGAAATGATGTTGTAGCAT 1446
XX
seq_name: /SID56/gcgdata/geneseq/NA1997.DAT:V74477
seq_documentation_block:
ID V74477 standard: DNA, 9623 BP.
AC V74477:
XX

```

```

DT 16-MAR-1999 (first entry)
XX
XX Staphylococcus aureus contig SEQ ID #166.
DE
XX Computer readable medium; vaccine; S.aureus infection; immunodetection;
XX celluilitis; eyelid infection; food poisoning; osteomyelitis; therapy;
XX skin infection; surgical wound infection; scalded skin syndrome;
XX toxic shock syndrome; ds.
XX
XX Staphylococcus aureus.
OS
XX
XX Key Location/Qualifiers
FH
XX misc_feature 361..420
FT
XX /tag= a
FT "these bases represent a line of missing text in
FT the sequence listing in the specification. They
FT are included to maintain the nucleotide numbering
FT given in the specification for this DNA sequence"
FT
XX misc_feature 2161..2220
FT
XX /tag= b
FT "these bases represent a line of missing text in
FT the sequence listing in the specification. They
FT are included to maintain the nucleotide numbering
FT given in the specification for this DNA sequence"
FT
XX misc_feature 3961..4020
FT
XX /tag= c
FT "these bases represent a line of missing text in
FT the sequence listing in the specification. They
FT are included to maintain the nucleotide numbering
FT given in the specification for this DNA sequence"
FT
XX misc_feature 5761..5820
FT
XX /tag= d
FT "these bases represent a line of missing text in
FT the sequence listing in the specification. They
FT are included to maintain the nucleotide numbering
FT given in the specification for this DNA sequence"
FT
XX misc_feature 7561..7620
FT
XX /tag= e
FT "these bases represent a line of missing text in
FT the sequence listing in the specification. They
FT are included to maintain the nucleotide numbering
FT given in the specification for this DNA sequence"
FT
XX misc_feature 9361..9420
FT
XX /tag= f
FT "these bases represent a line of missing text in
FT the sequence listing in the specification. They
FT are included to maintain the nucleotide numbering
FT given in the specification for this DNA sequence"
FT
XX EP786519-A2.
XX
XX 30-JUL-1997.
XX
XX 07-JAN-1997; 97EP-0100117.
XX
XX 05-JAN-1996; 96US-0009861.
XX
XX (HUMA-) HUMAN GENOME SCI INC.
XX
XX Barash SC, Choi GH, Dillon PJ, Fannon MR, Kunsch CA:
XX Rosen CA;
XX
XX WPI; 1997-374922/35.
XX
XX Polynucleotide(s) and proteins derived from Staphylococcus aureus -
XX stored on computer readable medium and used in the production of
XX anti-S.aureus vaccines
XX
XX Claim 1; Page 816-822; 3271pp; English.
XX
XX This sequence represents one of 5191 Staphylococcus aureus DNA sequences
XX of the invention. The DNA sequences are recorded on a computer accessible
XX medium, preferably selected from a floppy or hard disk, random access
XX

```


alignment_scores:

Quality: 619.00 Length: 299
Ratio: 2.906 Gaps: 4
Percent Similarity: 71.237 Percent Identity: 39.465

alignment_block:

US-09-462-846-2 x V74326/rev ..

Align seg 1/1 to reverse of: V74326 from: 1 to: 3759

```

3 ThrGluProLeuPhePheLysProValPheLysGluArgIleTrpGly 19
   ||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
972 ACAATGCCATTAATTTTACACCAATTTTAAACAAATTAATGCGCG 923
   ||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
19 yThrAlaLeuAlaAspPheGlyTrpThrIleProSerGlnArgThr.Gly 35
   ||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
922 TCACGCTCTAAGTAGAGTTGATATCAATTAAGCAATGATACACTGGGG 873
   ||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
36 GluGlySerPheAlaPheAlaAlaHisGlnAsnGlyLysSerValValGlnAs 52
   ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
872 GAAATGTGGGTGTGTGCACACATCCAAATGTGACAGCGAGATTTTAA 823
   ||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
52 nglyMetTrpLysGlyPheThrLeuSerGluLeuTrpGluHisHisArgH 69
   ||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
822 TGGACCATATCAAGTCTCAAAACATTAGACCGTATTGTCAGAACATCGTG 773
   ||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
69 IsteuPheGlyGlnLeuGluGlyAspArgPheProLeuLeuThrIle 85
   ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
772 AATTGTTGGTGTATTCACAGCAAGATTTCGCTTACTAATAAATA 723
   ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
86 LeuAspAlaAspGlnAspLeuSerValGlnValHisProAsnAspGlu 102
   ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
722 GTGGATGCAAGAGAAATCTTCTATTCATGTGCACCCGATTAATCTTA 673
   ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
102 rAlaAsnIleHisGlnAsnGlyLeuGlyLysThrGluCysTrpTrpTr 119
   ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
672 TCGCTATGAGCATGAACACGGGCAATATGCAAAATCTGAATGTGTATA 623
   ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
119 IetIleAspCysGlnLysAspAlaGluIleIleIleTrpGlyHisAsnAla 135
   ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
622 TTATTAATGTCAGAAAGATGCGAAATAGTTATAGGACATTAGACAGAG 573
   ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
136 ThrLysGluGluLeuThrThrMetIleGluArgGluTrpAspGlu 152
   ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
572 TCTAGAGAAAGAGTTCGAATCATGTTCACACGAGACATAGATCGAT 523
   ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
152 uLeuArgArgValLysValLysProGlyAspPhePheTrpValProSerG 169
   ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
522 ACTTGATATATTAAGTAAACCTGCGAATTCATTTTATNNNNNNNN 473
   ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
169 LythrValHisAlaIleGlyLysGlyIleLeuAlaLeuGluThrGln 185
   ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
472 NNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNN 423
   ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
186 AsnSerAspThrThrArgLeuTrpAspTrpAspArgLysAspAla 202
   ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
422 NNTCTAGACATTTACATATGACTTTATGATTTCAATCGCAAGATATCA 373
   ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
202 uGlyLysLeuArgGluLeuHisLeuLysLysSerIleGluValIleGlu 218
   ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
372 ATATATGATAGACCGTAAATATTCAAAAAGCTTATAGCGTTATTCAT 323
   ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
219 .....ValProSerIle..ProGluArgHisThrValHisHis 230
   ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
322 ACAATGCCACATTACTAATATTTGCGCGAA.....AGC 288
   ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
231 GluGlnIleGluAspLeuLeuThrThrThrLeuIleGluCysAlaTrp 247
   ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
287 GAAATTTATGAAACCATTAAGTCTACACACATTTGTTCAATGATTTCT 238
   ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
247 eSerValGlyLysTrpAsnLeuSerGlySerAlaSerLeuLysGlnGln 264

```

seq_name: /SID56/gcgdata/geneseq/geneseqn/NA1999.DAT.X20249

seq_documentation_block:

ID X20249 standard; DNA; 116277 BP.

XX X20249;

DT 04-MAY-1999 (first entry)

XX Borrelia burgdorferi polynucleotide sequence #2.

KW Borrelia burgdorferi; spirochete; bacterium; pathogen; Lyme disease;

KW epidemic relapsing fever; endemic relapsing fever; Lyme borreliosis;

KW infection; diagnosis; characterisation; detection; ds.

OS Borrelia burgdorferi.

PN W09858943-A1.

XX 30-DEC-1998.

PF 18-JUN-1998; 98MO-US12764.

XX 03-SEP-1997; 97US-0057483.

PR 20-JUN-1997; 97US-0050359.

PR 22-JUL-1997; 97US-0053344.

XX 22-JUL-1997; 97US-0053377.

PA (HUMA-) HUMAN GENOME SCI INC.

PA (MED1-) MEDIMUNE INC.

PI Clayton R, Dougherty BA, Fraser C, Lathigra R, Smith HO;

PI White OR;

DR WPI; 1999-081217/07.

XX New isolated Borrelia burgdorferi nucleic acids - used to develop

PT products for the detection, diagnosis, characterisation, prevention

PT and therapy of infections, particularly Lyme disease

XX Claim 1; Page 672-737; 1128pp; English.

PS X20248 to X20402 represent polynucleotide sequences isolated from

CC Borrelia burgdorferi (Bb). Products derived from Bb can be used for

CC the detection, diagnosis, characterisation, prevention and therapy of

CC Bb infections, e.g. Lyme disease. They can also be used for the

CC production of biosynthetic products, e.g. enzymes. Borrelia belongs

CC to a family of motile, spiral-shaped bacteria called Spirochetes.

CC Spirochetes are pathogenic in humans and Borrelia causes epidemic and

CC endemic relapsing fever, and Lyme borreliosis, more commonly known as

XX Lyme disease.

XX Sequence 116277 BP; 42656 A; 19868 C; 14490 G; 39250 T; 13 other:

alignment_scores:

Quality: 166.00 Length: 406
Ratio: 0.883 Gaps: 22
Percent Similarity: 46.305 Percent Identity: 23.645

alignment_block:

US-09-462-846-2 x X20249/rev ..

Align seg 1/1 to reverse of: X20249 from: 1 to: 116277

```

1 MethTrhGluProLeuPhePhe.....LysProValPheLysGluAr 15
  |||.....|||.....|||.....|||.....|||.....|||.....
63923 ATGATTAATTAAGATTAATATTTTAAATGAAATAATATTAAGATA 63874
15 g1eTrpGlyThrAlaLeuAlaAsp.....PheGlyTyrThrIleP 30
  |||.....|||.....|||.....|||.....|||.....|||.....
63873 TGATTTGGCGGAATTAATTTATCCCAATCTTTGGGATAGATAG 63824
30 roserGlnArgThrGlyLcysTrpAlaPheAlaAlaHisGlnAsnGly 46
  |||.....|||.....|||.....|||.....|||.....|||.....
63823 ATGGAAGCCCAAGCGTGAATGTGG...CTTGAGCACACAAAGACATTT 63777
47 GlnSerValAlaGlnAsnGlyMetTyrLysGly.....PheThrIleu 61
  |||.....|||.....|||.....|||.....|||.....|||.....
63776 TCTAGTAAAGATT.....TGTATTAATAATGATATGTGCTTTTAA 63736
61 rGlnLeuTrpGlnHisArgHisLeuPheGlyGlnLeuGlnLysAsp 78
  |||.....|||.....|||.....|||.....|||.....|||.....
63735 CGATTTTAAAGATCATATAAGACCTTTTAGCGTGT....AATGACG 63692
78 rGpHeProLeuLeuThrLysIleLeuAspAlaAspGlnAspLeuSerVal 94
  |||.....|||.....|||.....|||.....|||.....|||.....
63691 AATTCCTTTTCTTACGATATGTCTGCAATAAGCCCTGTCTATT 63642
95 GlnValHisProAsnAspGlu.....TyrAlaAsnIleH 106
  |||.....|||.....|||.....|||.....|||.....|||.....
63641 CAATATCTCTCTTAAGATATTGCTTAAAGCGTATGATGAGGAA 63592
106 sGlnAsnGly.....GlnL 111
  |||.....|||.....|||.....|||.....|||.....|||.....
63591 TAATAAAGGATAGACATTAATGATCCCAAAAGACATACAAACAAA 63542
111 euGlyLysThrGlyCysTrpTyrIleIle..... 120
  |||.....|||.....|||.....|||.....|||.....|||.....
63541 ACCCCAAATTAAGACTTATTTATCTCTAGTATTTTATGCTTTAA 63492
121 .....AspCysGlnLysAspAlaGluIleIle..... 129
  |||.....|||.....|||.....|||.....|||.....|||.....
63491 GCGTTTACCCTTAGATGACATTAATAAAATTTATGAAATTCGGAATT 63442
130 .....TyrGlyHisAsnAlaThrThrLysGlu..... 138
  |||.....|||.....|||.....|||.....|||.....|||.....
63441 AAATTCGACTTTCATTCACATTAAGATTTGTAAAGACATTTTGATT 63392
139 .....GlnLeuThrThrMetIleGlu..... 145
  |||.....|||.....|||.....|||.....|||.....|||.....
63391 TACAATGTATGAACCTTGAGAAATTTATGAAAAAATTTAAAAAATTG 63342
146 .....ArgGlyGluTrp...AspGlnLeu..... 152
  |||.....|||.....|||.....|||.....|||.....|||.....
63341 GATCTTATTTGATTAATTTAGGGCTATTGGTTTATGCAATTTTCAATAT 63292
153 .....LeuArgArgValL 157
  |||.....|||.....|||.....|||.....|||.....|||.....
63291 TTATGTATATAGATGAGCCCTTTTGTATTTTATAGTATGAATTTTAA 63242
157 ysValLysProGlyAspPhePheTyrValProSerGlyThrValHisAla 173
  |||.....|||.....|||.....|||.....|||.....|||.....
63241 AACTAATAACCGAGAAAGTTGTTATACAAATAGTCAGAGAGCTGCATGCA 63192
174 IleGlyLysGlyIleLeuAlaLeuGluThrGlnGlnAsnSerAspThrTh 190
  |||.....|||.....|||.....|||.....|||.....|||.....
63191 TATCTTAAGGAGAT...TGCATTGACCTTATGACCAATTCGACAAATGT 63145
190 rTyrArg.....LeuTyrAspTyrAspArgLysAsp..... 200
  |||.....|||.....|||.....|||.....|||.....|||.....
63144 TATTAGGCGCTGGCTTACTACAAAGTATATTGATAAAGACAGATGTAA 63095
201 .....AlaGlnGlyLysLeuArgGlnLeuHis.....Leu 210
  |||.....|||.....|||.....|||.....|||.....|||.....

```

```

63094 GAGTTGTCATATTGAGAGAAAGTATCATTTTAAATCCGATTTT 63045
211 LysLysSerIleGlnValIleGlnValProser..... 221
  |||.....|||.....|||.....|||.....|||.....|||.....
63044 CAGGATTAATTTTAGCGTATTTAGACTTCCAAATTAATTTAAATTGAT 62995
222 .....IleProGluArgHisThrValHisHisGlnGlnIleGlnVal 235
  |||.....|||.....|||.....|||.....|||.....|||.....
62994 TCATAAAAAAATTAATGAGACACTTGTATTAATGAAATAGTCAATGG 62945
235 sPheLeuThrThrThrLeuIleGlnCysAlaTyrPheSerValGlyLys 251
  |||.....|||.....|||.....|||.....|||.....|||.....
62944 TCTTCTAGTTTAAAT.....GGTGC 62922
252 TrpAsnLeuSerGlySerAlaSerLeuLysGlnGlnLysProPheLeu 268
  |||.....|||.....|||.....|||.....|||.....|||.....
62921 GTGAGTATTAATTAATCTTAATCTTAAGAAA..... 62889
268 uIleSerValIleGlnGlyLysArgMetIleSerGlyGluTyrValT 285
  |||.....|||.....|||.....|||.....|||.....|||.....
62888 .....GGTGAAGCATAT 62876
285 yTrProPheLysGlyAspHisMetLeuLeuProTyrGlyLeuGlyGlu 301
  |||.....|||.....|||.....|||.....|||.....|||.....
62875 TTATAGTAAATAAGCAGAAACTTGTATTGAT...GGGACGCGCGAA 62829
302 PheLysLeuGlnGlyTyr 307
  |||.....|||.....|||.....|||.....|||.....|||.....
62828 GCTTTTATTTGCTGCTTTT 62811

```

seq.name: /SID56/gcdata/geneseq/geneseqn/NA1999.DAT:X20248

seq_documentation_block:

ID X20248 standard; DNA: 910715 BP.

XX X20248;

DT 04-MAY-1999 (first entry)

DE Borrelia burgdorferi polynucleotide sequence #1.

KM Borrelia burgdorferi; spirochete; bacterium; pathogen; Lyme disease;
 KW epidemic relapsing fever; endemic relapsing fever; Lyme borreliosis;
 OS infection; diagnosis; characterisation; detection; ds.

OS Borrelia burgdorferi.

PN W09858943-A1.

PD 30-DEC-1998.

PE 18-JUN-1998; 98WO-US12764.

PR 03-SEP-1997; 97US-0057483.

PR 20-JUN-1997; 97US-0050359.

PR 22-JUL-1997; 97US-0053344.

PR 22-JUL-1997; 97US-0053377.

PA (HUMA-) HUMAN GENOME SCI INC.

PI (MEDT-) MEDIMUNE INC.

PI Clayton R, Dougherty BA, Fraser C, Iathigra R, Smith HO;

DR WPL: 1999-081217/07.

PT New isolated Borrelia burgdorferi nucleic acids - used to develop

PT products for the detection, diagnosis, characterisation, prevention

PT and therapy of infections, particularly Lyme disease

PS Claim 1; Page 157-671; 1128pp; English.

XX X20248 to X20402 represent polynucleotide sequences isolated from

CC Borrelia burgdorferi (Bb). Products derived from Bb can be used for
 CC the detection, diagnosis, characterisation, prevention and therapy of
 CC Bb infections, e.g. Lyme disease. They can also be used for the
 CC production of biosynthetic products, e.g. enzymes. Borrelia belongs
 CC to a family of motile, spiral-shaped bacteria called Spirochetes.
 CC Spirochetes are pathogenic in humans and Borrelia causes epidemic and
 CC endemic relapsing fever, and Lyme borreliosis, more commonly known as
 CC Lyme disease.

XX Sequence 910715 BP; 327171 A; 129646 C; 130753 G; 323091 T; 54 other;

alignment_scores: Quality: 166.00 Length: 406
 Ratio: 0.863 Gaps: 22

Percent Similarity: 46.305 Percent Identity: 23.645

alignment_block:

US-09-462-846-2 x X20248 ..

Align seg 1/1 to: X20248 from: 1 to: 910715

```

1 MetThrThrGluProLeuPhePhe.....LysProValPheLysGluLeu 15
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
491204 ATGAATATGACATATATATTTTAAATGAAAAAATAATATTAAGAAATA 491253
15 g11eTPGlyGlyThrAlaLeuAlaSP.....PheGlyTyrThrIleP 30
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
491254 TGATGGGGGGGGAATTAATTTATCCCAATCTTTGGGATAGAAATGG 491303
30 roserGlnArgThrGlyLucyStryPheAlaAlaHisGlnAsnGly 46
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
491304 ATGGAAGCCCAAGCGCTGAATGTGG...CTTGAGCAGCAGACAAATTT 491350
47 GlnSerValValGlnAsnGlyMetTyrLysGly.....PheThrLeu 61
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
491351 TCATGTAAAGTT.....TTGTATAAAAAATGAATATGTCTTTAAG 491391
61 rGluLeuTrrGluHisHisArgHisLeuPheGlyGlnLeuGluGlyAsp 78
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
491392 CGATTTTATAGAAATCATTAAGCGCTTTAGCTGT.....AATGACG 491435
78 rGpPheProLeuLeuThrLysIleLeuAspAlaAspGlnAspLeuSerVal 94
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
491436 AATTTCCTTTTTCCTTAGATATGTCTGCAATAAAGCCCGTGTATTT 491485
95 GlnValHisProAsnAspGlu.....TyrAlaAsnIleH 106
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
491486 CAAATTCATCCTCTTAAGATATTGCCTTAAAGGATATGATCAGAGAA 491535
106 sGluAsnGly.....GluL 111
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
491536 TAAATTAAGGGATAGACATTAATGATCCCAAAAGACATACAAAGACAAA 491585
111 euGlyLysThrGlyCysTrrPyrIleIle..... 120
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
491586 ACCCCAAAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTA 491635
121 .....AspCysGlnLysAspAlaGluIleIle..... 129
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
491636 GCGTTTATACCTTAGATAGATTAATAAAATTAATTAATTAATTAATTA 491685
130 .....TyrGlyHisAsnAlaThrThrLysGlu..... 138
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
491686 AAATTTGCATCTTCATACATTAAGATTTTGTAAAGACTTTTGTGATT 491735
139 .....GluLeuThrMetIleGlu..... 145
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
491736 TACAAATGTATGAAGCTGAGAAATTAATTAATTAATTAATTAATTAAT 491785
146 .....ArgGlyGluTrp...AspGluLeu..... 152
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
491786 GATCTATATGATTAATTTTAAAGGCGTATTGGTTTAAATTAATTAATTA 491835

```

```

153 .....LeuArgArgVal 157
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
491836 TTATGGTATAGATGTGGCCCTTTGGTATTTTATAGATATGAATTTTAA 491885
157 ysValLysProGlyAspPhePheTyrValProSerGlyThrValHisAla 173
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
491886 AACTTAAACACAGAGAAAGTGTATACAAATAGTCAGAGAGTCAGCA 491935
174 IleGlyLysGlyIleLeuAlaLeuGlnThrGlnGlnAsnSerAspThr 190
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
491936 TATCTTAAGGAGAT...TGCATTGAGCTTATGACCAATTCGCAATGT 491982
190 TTYArg.....LeuTyrAspTyrAspArgLysAsp..... 200
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
491983 TATTAGGCGCTGGCTTACTACAAAGTATATGATTAAGACAGATGTAA 492032
201 .....AlaGluGlyLysLeuArgGluLeuHis.....Leu 210
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
492033 GAGTTGTCATATTGAGGAGAAAGTATATCATTTTAAATCCGATTTT 492082
211 LysLysSerIleGluValIleGluValProSer..... 221
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
492083 CAAGATTAATTTTAGCGTATTTAGACTCCAAATACATAATTGAAATTTGAT 492132
222 .....IleProGluArgHisThrValHisGlnGluIleGlu 235
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
492133 TCAAAAAAAATAATGAGAACATTTGTATTAATTAATGAATGTGCAATG 492182
235 sPheLeuLeuThrThrThrLeuIleGluCysAlaTyrPheSerValGlyLys 251
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
492183 TCTCTAGTATTAAT.....GGGTGC 492205
252 TrpAsnLeuSerGlySerAlaSerLeuLysGlnGlnLysProPheLeu 268
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
492206 GTGAGATTAATAATAATCTTAATCTTAAGAA..... 492238
268 vIleSerValIleGluGlyGlyArgMetIleSerGlyGlyValT 285
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
492239 .....GGTGAAGACATAT 492251
285 yTrpPheLysLysGlyAspHisMetLeuLeuProTyrGlyLeuGlyGlu 301
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
492252 TTATAGCTAAAAAGCAGAAACTGTTTATGAT...GGGAGCGCGAA 492298
302 PheLysLeuGluGlyTyr 307
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
492299 GCTTTATTGTGTGTTT 492316
seq_name: /SID56/gcdata/geneseq/geneseqn/NN2000.DAT:A54211
seq_documentation_block:
ID A54211 standard; DNA: 6111 BP.
XX
XX A54211;
XX
XX 26-FEB-2001 (first entry)
XX
XX Transformation vector pCIB9818.
XX
XX Tricothecene resistance; resistant; crop protection; mycotoxin;
XX fungus; wheat; maize; barley; rice; heterologous gene;
XX transformation; Fusarium; ds.
XX
XX Synthetic.
XX
XX WO2000060061-A2.
XX
XX 12-OCT-2000.
XX
XX 29-MAR-2000; 2000WO-EP02769.
XX
XX 31-MAR-1999; 99US-0282995.

```



```

XX plasmid pZU634 for Potato virus Y (PVY) resistance in tomato.
DE
XX
XX Potyvirus: PVY: virus resistant; virus tolerant;
KW tomato; coat protein; CP; plasmid pZU634; ds.
XX
XX Chimeric - Potato virus Y.
OS Chimeric - Bacteria.
OS Chimeric - Agrobacterium sp.
OS Chimeric - Unidentified.
XX
FH Key
FT Location/Qualifiers
FT 55..1786
FT /*tag= a
FT /*note= "Ubiquitin3 (Ubi3) promoter plus leader
FT intron"
FT 1790..3847
FT /*tag= b
FT /*note= "Potato virus Y (PVY) inverted repeat region"
FT 1790..2574
FT /*tag= c
FT /*rpt_type= INVERTED
FT /*note= "Corresponds to PVY coat protein (CP) region"
FT 2595..3036
FT /*tag= d
FT /*label= Actin2_intron1
FT complement (3057..3847)
FT /*tag= e
FT /*rpt_type= INVERTED
FT /*note= "Corresponds to PVY CP region"
FT 3855..4124
FT /*tag= f
FT /*label= NOS
FT /*note= "Agrobacterium nopaline synthase terminator"
FT 4193..5364
FT /*tag= g
FT /*label= SMAS_promoter
FT 5386..6561
FT /*tag= h
FT /*product= "Phosphomannose isomerase A"
FT 6623..6897
FT /*tag= i
FT /*label= NOS
FT /*note= "Agrobacterium nopaline synthase terminator"
FT 6949..7080
FT /*tag= j
FT /*label= Nopaline_left_border_fragment
FT 7359..8147
FT /*tag= k
FT /*product= "Bacterial spectinomycin-detoxifying enzyme,
FT aminoglycoside-3'-adenylyltransferase"
FT 8157..10244
FT /*tag= l
FT /*label= pVSI_ORI
FT 10245..11179
FT /*tag= m
FT /*label= pUC19_ORI
FT 11195..11450
FT /*tag= n
FT /*label= Nopaline_right_border_fragment
XX
XX WO200068374-A1.
XX
XX 16-NOV-2000.
XX
XX 08-MAY-2000; 2000WO-EP04117.
XX
XX 10-MAY-1999; 99US-0309038.
XX
XX (NOVS ) NOVARTIS AG.
XX (NOVS ) NOVARTIS-ERFINDUNGEN VERW GES MBH.
XX
XX Haefetz PB, Patton DA, Levin JZ, Que Q, De Haan PT, Gielens JJL,
PI

```

```

XX
XX WPI: 2001-007392/01.
XX
XX Altering the expression of a viral genome, to obtain cells, plants or
XX animals that are resistant or tolerant to viruses, comprises using
XX sense and antisense RNA fragments that form double-stranded RNA
XX molecule -
XX
XX Claim 14; Page 67-71; 75pp; English.
XX
XX The present invention relates to methods of altering the expression of
XX a viral genome, comprising introducing into a cell, a first DNA sequence
XX capable of expressing a sense RNA fragment of the viral genome and a
XX second DNA sequence capable of expressing an antisense RNA fragment of
XX the viral genome. The sense and antisense RNA fragments are capable of
XX forming double-stranded RNA. The method is useful for rendering cells,
XX preferably plant cells, animals or their progenies resistant or tolerant
XX to viruses such as tospoviruses, potyviruses, potexviruses,
XX tobamoviruses, luteoviruses, cucumoviruses, bromoviruses,
XX closteroviruses, tombusviruses and furoviruses.
XX The present sequence is plasmid pZU634, a chimeric gene cassette
XX comprising the coding sequence for coat protein (CP) from potato virus
XX Y (PVY). This plasmid comprises a ubiquitin3 (Ubi3) promoter plus
XX leader intron, SMAS promoter, nopaline synthase (NOS) terminator,
XX actin2 intron, pVSI ORI, pUC19 ORI and coding sequences for
XX phosphomannose isomerase A and bacterial spectinomycin-detoxifying
XX enzyme, aminoglycoside-3'-adenylyltransferase. This plasmid is used to
XX confer resistance to PVY in tomato.
XX
XX Sequence 11461 BP; 3128 A; 2581 C; 2645 G; 3107 T; 0 other:
XX
XX

```

```

alignment_scores:
Quality: 145.50 Length: 398
Ratio: 0.841 Gaps: 17
Percent Similarity: 43.467 Percent Identity: 21.357
alignment_block:
US-09-462-846-2 x D02175 ..

```

Align seg 1/1 to: D02175 from: 1 to: 11461

```

10 ProValPheLysGluArgLe.....TTPGlyGI 19
   |||:|||||:|||||
5380 CCGATCATGCAAAATGCTATTAAGTCAAGTCAAAATGCTGCTGGGGAG 5429
   |||:|||||:|||||
19 Y...ThrAlaLeuAlaasp...PheGlyTYrThrIlePro...serGln 33
   : |||:|||||:|||||:|||||
5430 CAAAAGCGGCTTGACTGAACCTTTATGATGGAATAATCCGTCACAGCCAGC 5479
   : |||:|||||:|||||:|||||
33 rGThrGlyGluCysTrpAlaPheAlaAlaHisGlnAsnGlyGlnSerVal 49
   : |||:|||||:|||||:|||||
5480 CGATGGCGGAGCTGTGG...ATGGCGGCACATCCGAAAGCAGTTCAAGCA 5526
   : |||:|||||:|||||:|||||
50 ValGlnAsnGlyMetTyrLysGlyPheThrIleuSerGluIleuTrpGlnI 66
   : |||:|||||:|||||:|||||
5527 GTGCAGAAATGCCCGCGAGATATCGTTTCACATCGCTGATGATTAAGAG 5576
   : |||:|||||:|||||:|||||
66 sHisArgHis...LeuPheGlyGlnLeuGluGlyAspArgPhe..... 79
   : |||:|||||:|||||:|||||
5577 TGATTAATGACTCTGCTGGAGAGGCCCTTCCCAACGCTTTGGCGAAGC 5626
   : |||:|||||:|||||:|||||
80 ..ProLeuLeuThrIlySileuAspAlaAspGlnAspLeuSerValGln 95
   : |||:|||||:|||||:|||||
5627 TGCCCTTTCCTGTTCAAAAGATATATGCGCAGCAGCAGCCACTCTCCATTG 5676
   : |||:|||||:|||||:|||||
96 ValHisProAsnAspGluTyrAlaAsnIleHisGluAsnGlyGluLeuGI 112
   : |||:|||||:|||||:|||||
5677 GTTCATCAACAAC.....AAACACAAATTCTGAATAATCG 5708
   : |||:|||||:|||||:|||||
112 Y.....
5709 TTTTGGCAAGAAATGCGGACAGTATCCGATGGATGGCGCGGAGCGTA 5758

```

```

113 .....lysthrGluCysTrpTyrIleIle..... 120
    ||| |||
5759 ACTATAAGATCCCTAACACACAGCCGAGCTGTTTGGCGTGAGCCCT 5808
120 ..... 120
5809 TTCCTTGCGATGAACGGCTTCGTGAATTTCCGAGATTGTCCTACT 5858
120 ..... 120
5859 CCAGCCGCTCCAGTGCACATCCGGCATTCCTCACTTTTACACAGC 5908
121 ..AspCysGlnLysAspAlaGluIleIleTyrGlnHisAlaIleThr 136
5909 CTGATGCCGAACTTTAAGCAACTGTTCCGACCGCTGTAATATGACAG 5958
137 LysGlnGluLeuThrThiMetIle..... 144
    |||||
5959 GGTAAAGAAAATCCCGCGCGCTGGCATTTTAAATCGCCCTCGATAG 6008
145 ..GluArgGlyGlu...TrpAspGlu..... 151
    ::::: |||||
6009 CCAGCAGGCTGAACCGTGCACAAAGCATTCGTTAATTTCTGAATTTACC 6058
152 .....LeuLeuArgValLysVal 158
    |||||
6059 CGAAGACAGCGCTCTCTCCCGCTATTCGTGAATGTGGAAATAG 6108
159 LysProGlnAspPhePheTyrValProSerGlyThrValHisAlaIleG 175
    |||||
6109 AACCTGGCGAAGCGATGTCCTGCTGCGTGAACACCGCAGCTTACCT 6158
175 ylyGlyLysLeuAlaGlnThrGlnGlnAsnSerAspThrTyr 192
    ::::: |||||
6159 GCAGGGCGCTG...GCCTGGAGATGAGCAAACTCCGATAACGCTG... 6201
192 rGluTyrAspTyrAspArgLysAspAlaGlnLysLeuArgGlnLeu 208
    |||
6202 .....CTG 6204
209 HisLeuLysLysSerIleGlnValIleGlnValProSerIleProGlu 225
    ::::: |||||
6205 CGTGGCGCTGACGCCCTAAATACATTGATATTCGGAACTGGTGCCA 6254
225 gHisThrValHisLysGlnIleGlnAspLeuLeuThrThrLeu 242
    ::::: |||||
6255 TGTGAATTCGAAAGCAAAACCGCTAACAGTTGTTGACCCAGCCGCTGA 6304
242 leGluCysAlaTyr.....PheSerVal 249
    ::::: |||||
6305 AACAGAGTGCAGAACTGACTCCCGATTCACATGATGATTTTGCCCTTC 6354
250 GlyLysTrpAsnLeuSerGlySerAlaSerLeuLysGlnLysProP 266
    ::::: |||||
6355 TCGCGCATGACCTTAGTGAATAAGAAACACCATTAGCAGCAGAGTGC 6404
266 leuLeuLeuSerValIleGlnGlnGlnGlnArgMetIleSerGlyL 283
    ::::: |||||
6405 CGCCATTTTGTTCGCGTGAAGCGATGCAAGCTTGGAAGAGTTCTC 6454
283 yValTyrProPheLysLysGlyAspHisMetLeuLeuPro 296
    ::::: |||||
6455 AGCAGTTACAGCTTAACCGGCTGAATCAGCGTTTATTTGCCG 6496
seq_name: /SIDS6/gcgdata/geneseq/geneseqn/NA2001.DAT:DD2174
seq_documentation_block:
ID DD2174 standard; DNA; 12766 BP.
XX
AC DD2174;
XX
DT 28-MAR-2001 (first entry)

```

```

XX Plasmid pZU623 for ZYMV and PRSV resistance in melon.
DE
XX Potyvirus: ZYMV; PRSV: virus resistant; virus tolerant;
KW melon; coat protein; CP; plasmid pZU623; ds.
KW
XX Chimeric - Zucchini yellow mosaic virus.
OS
XX Chimeric - Papaya ringspot virus.
OS
XX Chimeric - Bacteria.
OS
XX Chimeric - Agrobacterium sp.
OS
XX Chimeric - Unidentified.
XX
PH Key
FT Location/Qualifiers
FT 55..1781
FT /tag= a
FT /note= "Ubiquitin3 (Ub13) promoter plus leader
FT intron"
FT 1790..5143
FT /tag= b
FT /note= "PRSV-ZYMV inverted repeat fragment"
FT 1790..2430
FT /tag= c
FT /rpt_type= INVERTED
FT /note= "Corresponds to PRSV coat protein (CP) region"
FT 2511..3111
FT /tag= d
FT /rpt_type= INVERTED
FT /note= "Corresponds to ZYMV CP region"
FT 3245..3686
FT /tag= e
FT /label= Actin2_intronL
FT complement (3822..4422)
FT /tag= f
FT /rpt_type= INVERTED
FT /note= "Corresponds to ZYMV CP region"
FT complement (4503..5143)
FT /tag= g
FT /note= "PRSV CP region"
FT complement (5160..5429)
FT /tag= h
FT /note= "Agrobacterium nopaline synthase terminator"
FT 5498..6669
FT /tag= i
FT /label= SMS_Promoter
FT 6691..7866
FT /tag= j
FT /product= "Phosphomannose isomerase A"
FT 7928..8202
FT /tag= k
FT /label= NOS
FT /note= "Agrobacterium nopaline synthase terminator"
FT 8254..8385
FT /tag= l
FT /label= NOS_Promoter
FT 8664..9452
FT /tag= m
FT /product= "Bacterial spectinomycin-detoxifying enzyme,
FT aminoglycoside-3'-adenylyltransferase"
FT 9462..11549
FT /tag= n
FT /label= pvs1_ORI
FT 11550..12484
FT /tag= o
FT /label= pUC19_ORI
FT 12500..12755
FT /tag= p
FT /label= Nopaline_right_border_fragment
PD 16-NOV-2000.

```



```

XX 08-MAY-2000; 2000MO-BP04117.
PF
XX 10-MAY-1999; 99US-0309038.
PR
XX (NOVS ) NOVARTIS AG.
PA (NOVS ) NOVARTIS-ERFINDUNGEN VERM GES MBH.
XX
PI Helfetz PB, Patton DA, Levin JZ, Que Q, De Haan PT, Gielen JTL;
DR WPI; 2001-007392/01.
XX
PT Altering the expression of a viral genome, to obtain cells, plants or
PT animals that are resistant or tolerant to viruses, comprises using
PT sense and antisense RNA fragments that form double-stranded RNA
PT molecule -
XX
PS Claim 14; Page 62-65; 75pp; English.
XX
CC The present invention relates to methods of altering the expression of
CC a viral genome, comprising introducing into a cell, a first DNA sequence
CC capable of expressing a sense RNA fragment of the viral genome and a
CC second DNA sequence capable of expressing an antisense RNA fragment of
CC the viral genome. The sense and antisense RNA fragments are capable of
CC forming double-stranded RNA. The method is useful for rendering cells,
CC preferably plant cells, animals or their progenies resistant or tolerant
CC to viruses such as tospoviruses, potyviruses, potexviruses,
CC tobamoviruses, luteoviruses, cucumoviruses, bromoviruses,
CC closteroviruses, tombusviruses and furoviruses.
CC The present sequence is plasmid pZU623, a chimeric gene cassette
CC comprising the coding sequence for coat protein (CP) from Zucchini
CC yellow mosaic virus (ZYMV) and Papaya ringspot virus (PRSV). This
CC plasmid also comprises a ubiquitin3 (Ub3) promoter plus leader
CC intron, SMAS promoter, nopaline synthase (NOS) terminator, actin2
CC intron, pVSI ORI, pUC19 ORI and coding sequences for phosphomannose
CC isomerase A and bacterial spectinomycin-deoxylyzing enzyme,
CC aminoglycoside-3'-adenylyltransferase. This plasmid is used to confer
CC resistance to ZYMV and PRSV in melon.
XX
SO Sequence 12766 BP; 3472 A; 2890 C; 2955 G; 3449 T; 0 other:

alignment_scores:
Quality: 145.50 Length: 398
Ratio: 0.841 Gaps: 17
Percent Similarity: 43.467 Percent Identity: 21.357

alignment_block:
US-09-462-846-2 x D02174 ..

Align seg 1/1 to: D02174 from: 1 to: 12766

10 ProvalphelysgluArgle.....Trrglycl 19
|||||:|||||:|||||:|||||:|||||:
6685 CCGATCATGCAAAACATTAATCACTGACGCAAACTATGCCTGGGCGAG 6734
19 Y...Thralaleualaasp...PheglYtrThrlePro...Sergina 33
Y:|||||:|||||:|||||:|||||:|||||:
6735 CAAACGGCGTTGACTGAACCTTATGATGGAATAATCCGTCACACCGCAGC 6784
33 rgtHrGlyGluCysTrpAlaPheAlaHisGlnAsnGlyGlnSerVal 49
:::|||||:|||||:|||||:|||||:|||||:
6785 CGATGGCCGAGCTGTGG...ATGGGCGCACATCCGAAAGACAGTTCACGA 6831
50 ValGlnAsnGlyMetYrlysglyPheThrLeuSerGluLeuTrpGluHr 66
|||||:|||||:|||||:|||||:|||||:
6832 GTGCGAAGAACCGCCGAGATATCGTTTCACGCGCTGATGTGATGAGAG 6881
66 sHisArgHis...LeupheGlyGlnLeuGluGlyAspArgPhe..... 79
:::|||||:|||||:|||||:|||||:|||||:
6882 TGATTAATGACTGTGCTGGAGAGCGCTTCCCAAACGCTTTGGCGAGAC 6931
80 ..ProleuLeuThrlyslleuAspAlaAspGlnAspLeuSerValGln 95

```

```

6932 TGCTTTCCTGTTCAAGATATATGCGACAGACGACGCTTCATTCAG 6981
96 ValHisProAsnAspGluYrlyAlaAsnIleHisGlnAsnGlyGlnLeu 112
|||||:|||||:|||||:|||||:|||||:
6982 GTTCATTCGCAAC.....AAACACAAATTCGAAATTCG 7013
112 Y..... 112
7014 TTTTGCAGAAAGAAATGCCGAGTATCCGATGATGCCCGACGAGTA 7063
113 .....LysThrGluCysTrpYrlyIle..... 120
7064 ACTATAAGATCTTAACACACAGCGAGCTGTTTGGCGTACGCT 7113
120 ..... 120
7114 TTCTTGGCATGACGCGTTTCGTGAATTTCCGAGATGTCCTCCTACT 7163
120 ..... 120
7164 CCAGCCGTCGCGAGTGCACATCCGCGATGCTCACTTTTACACAGC 7213
121 ..AspCysGlnLysAspAlaGluIleIleYrGlyHisAsnAlaThrTr 136
|||||:|||||:|||||:|||||:|||||:
7214 CTGATGCCGAAGCTTTAAGCAACTGTCGACGCTGTGATATACGAG 7263
137 LysGluGluLeuThrMetIle..... 144
7264 GGTGTAAGAAAAATCCCGCGCGCGATTTTAAATCGGCCCTCGATAG 7313
145 ..GluArgGlyGlu...TrpAspGlu..... 151
|||||:|||||:|||||:|||||:|||||:
7314 CCAGCAGAGTGGAACCGTGCAGAACGATCGTTAATTTGATTTTAC 7363
152 .....LeuLeuArgArgValLysVal 158
|||||:|||||:|||||:|||||:|||||:
7364 CGGAAGACAGCGTCTGTTCTCCCGCTATTCGTGATGGGGAATGG 7413
159 LysProGlyAspPhePheYrValProSerGlyThrValHisAlaIle 175
|||||:|||||:|||||:|||||:|||||:
7414 AACCTGCGAAGCGATGTTCTGCTGCGTGAACACCGCAGCTTACCT 7463
175 YLysGlyIleLeuAlaLeuGluThrGlnGlnAsnSerAspThrThrTyra 192
|||||:|||||:|||||:|||||:|||||:
7464 GCAAGGCGTG...GCGGTGAAGTGAATGCAACTCGATACAGTG... 7506
192 rgleuTyAspTyAspArgLysAspAlaGluGlyLysLeuArgGluLeu 208
|||||:|||||:|||||:|||||:|||||:
7507 .....CTG 7509
209 HisLeuLysLysSerIleGluValIleGluValProSerIleProGluArg 225
:::|||||:|||||:|||||:|||||:|||||:
7510 CGTCCGGGTCTGACGCTTAATATCATTTATTCGGAACGCTTGCCAA 7559
225 gHisThrValHisHisGlnGlnIleGluAspLeuLeuThrThrThrLeu 242
:::|||||:|||||:|||||:|||||:|||||:
7560 TGTAATAATGCAAGCCAAACCGGCTAACAGTTGTGTACCCACCGGTGA 7609
242 legLucysAlaTyR.....PheSerVal 249
:::|||||:|||||:|||||:|||||:|||||:
7610 AACAAAGTGCAGAACTGCATTCGCCATTCAGTGAATGATTTTGCTTC 7659
250 GlyLysTrpAsnLeuSerGlySerAlaSerLeuLysGlnIleLysProp 266
:::|||||:|||||:|||||:|||||:|||||:
7660 TCGCTGATGACCTTAAGTGAATAAGAAACCAATTCGACGAGAGTGC 7709
266 eleuLeuIleSerValIleGluGlyGluGlyArgMetIleSerGlyGlu 283
|||||:|||||:|||||:|||||:|||||:|||||:
7710 CGCATTTTGTTCGCTGAGAGCGATGCAAGCTTGTGAAAGGTTCTC 7759
283 yrValTyRProPheLysLysGlyAsp HisMetLeuLeuPro 296
:::|||||:|||||:|||||:|||||:|||||:

```

```

7760 AGCATTACAGCTTAAACCGGTGTAATCAGCGTTATTGCGG 7801
seq_name: /SIDS6/gcgdata/geneseq/geneseqn/NA2000.DAT: A54213
seq_documentation_block:
ID   A54213 standard: DNA, 12949 BP.
XX
AC   A54213;
XX
DT   26-FEB-2001 (first entry)
XX
DE   Transformation vector pNOV1704.
XX
KW   Tricothecene resistance; resistant; crop protection; mycotoxin;
KM   fungus; wheat; maize; barley; rice; heterologous gene;
KW   transformation; Fusarium; ds.
XX
OS   Synthetic.
XX
PN   WO200060061-A2.
XX
PD   12-OCT-2000.
XX
PF   29-MAR-2000; 2000WO-EP02769.
XX
PR   31-MAR-1999; 990US-0282995.
PR   11-FEB-2000; 2000US-0502852.
XX
PA   (NOVS ) NOVARTIS AG.
PA   (NOVS ) NOVARTIS-ERFINDUNGEN VERW GES MBH.
XX
PI   Hohn TM, Peters C, Salmeron JM, Reed JN, Dawson JL;
DR   WPI; 2000-679374/66.
XX
PT   Plant cell for preventing mycotoxin contamination of wheat, maize,
PT   barley or rice plant, comprises heterologous polynucleotide encoding
PT   gene product expressed in cell, having tricothecene resistance
PT   activity
XX
PS   Disclosure: Page 58-62; 62pp; English.
XX
CC   A heterologous gene encoding a gene product which confers
CC   tricothecene resistance can be used to transform plant cells to
CC   make them resistant to fungal infection. The transformation method is
CC   useful for preventing mycotoxin contamination of a plant,
CC   particularly a crop plant such as wheat, maize, barley or rice, and
CC   for reducing and/or preventing the growth of a fungus of the genus
CC   Fusarium that produces a tricothecene, preferably comprising a C-3
CC   hydroxyl group, by growing transformed crop plants in an area which
CC   is moderate to severe fungal infestation. Plasmid pNOV1704
CC   comprises the Zea mays ubiquitin promoter, the phosphate mannose
CC   isomerase selectable marker and the nopaline synthase
CC   termination sequence. pNOV1704 further comprises the Zea mays
CC   ubiquitin promoter operably linked to the tricothecene 3-O-acetyl
CC   transferase sequence given in GENESEQ record A54206 and the nos
CC   termination sequence.
XX
SQ   Sequence 12949 BP; 3194 A; 3146 C; 3143 G; 3466 T; 0 other;
XX
XX
alignment_scores:
    Quality: 145.50      Length: 398
    Ratio: 0.841        Gaps: 17
    Percent Similarity: 43.467    Percent Identity: 21.357
XX
alignment_block:
US-09-462-846-2 x A54213
..
Align seg 1/1 to: A54213 from: 1 to: 12949
10 ProValPheLysGluArgIle.....TyrGlyGlu 19
|||||:|||||:|||||

```

```

2011 CCGATCATGCAAAAACATCACTAATCACTGACGCAAAACATGCGTGGGCGAG 2060
19 Y...ThrAlaLeuAlaAsp...PheGlyTyrThrIlePro...SerGlna 33
: |||||:|||||:|||||:|||||
2061 CAAAACGGCGGTGACTGAACCTTATGATGTAAGGAAATCCGTCACCGCCAGC 2110
33 rGThrGlyGluCysTrpAlaPheAlaAlaHisGlnaSerGlna 49
: |||||: |||||: |||||: |||||: |||||: |||||: |||||: |||||
2111 CGATGGCCGAGCTGTGG...ATGGCGCACATCCGAAAGCAGTTCACGA 2157
50 ValGlnaSerGlyMetTyrLysGlyPheThrLeuSerGluLeuTrpGlu 66
|||||:|||||: |||||: |||||: |||||: |||||: |||||: |||||
2158 GTGCAGAAATGCCCGCGAGATATCGTTCACTGCGTGAATGATGATGACAG 2207
66 sHisArgHis...LeupheGlyGlnLeuGluGlyAspArgPhe..... 79
: |||||: |||||: |||||: |||||: |||||: |||||: |||||: |||||
2208 TGATAAATCGACTGTGCTGCGAGAGCGCGTTCGCAACGCTTTGGCGGAC 2257
80 ..ProLeuLeuThrLysIleLeuAspAlaAspGlnaAspLeuSerValGln 95
|||||:|||||: |||||: |||||: |||||: |||||: |||||: |||||
2258 TGCCCTTCTCTGTTCAAGATATATGCGACAGCACACCACTCTCCATTGAC 2307
96 ValHisProAsnAspGluTyrAlaAsnIleHisGlnaSerGlyGluLeuGlu 112
|||||:|||||: |||||: |||||: |||||: |||||: |||||: |||||
2308 GTTCATCCAAAC.....AAACAAATCTGAAATCGG 2339
112 Y..... 112
2340 TTTTGCCAAAGAAATGCGCGAGTATCCCGATGATGATCCCGCGAGCGTA 2389
113 .....LysThrGluCysTyrTyrIleIle..... 120
2390 ACTATAAAGATCTTAACCAACAGCGGAGCTGTGTTGGCGGTACGCGCT 2439
120 ..... 120
2440 TTCCTTGGAGTAAACGCGTTTCGTGAATTTCCGAGATGTGTCCTACT 2489
120 ..... 120
2490 CCAGCGCGTCCGAGGTGCACATCCGCGCATTTGCATCTTTTACACAGC 2539
121 ..AspCysGluLysAspAlaGluIleIleTyrGlyHisAsnAlaThrThr 136
|||||:|||||: |||||: |||||: |||||: |||||: |||||: |||||
2540 CCGATGCCGAAAGCTTAAAGCAACTGTTCGCCAGCTGTGAATATGACAG 2589
137 LysGluGluLeuThrThrMethIle..... 144
|||||: |||||: |||||: |||||: |||||: |||||: |||||: |||||
2590 GGTGAAGAAAATCCCGCGCGTGGCATTTTAAATCGCCCTCGATAG 2639
145 ..GluArgGlyGlu...TyrAspGlu..... 151
|||||: |||||: |||||: |||||: |||||: |||||: |||||: |||||
2640 CCAGAGAGGTGAACCGTGCACACGATTCGTTAATTCGTGAATTTTACC 2689
152 .....LeuLeuArgValValVal 158
|||||: |||||: |||||: |||||: |||||: |||||: |||||: |||||
2690 CGGAGACAGCGGTCTGTCTCCCGCTATTCGTGAAGTGTGTAATATG 2739
159 LysProGlyAspPhePheTyrValProSerGlyThrValHisAlaIleGlu 175
|||||: |||||: |||||: |||||: |||||: |||||: |||||: |||||
2740 AACCTGGCGAAGCGATTTCTCTGTGCTGAACACCGCACGCTTACCT 2789
175 yLysGlyIleLeuAlaLeuGluThrGlnGlnaSerAspThrThrTyr 192
|||||: |||||: |||||: |||||: |||||: |||||: |||||: |||||
2790 GCAAGCGCTG...GGCTGGAAGTATGCAAACTCCGATTAACGCTG... 2832
192 rGLeuTyrAspTyrAspArgLysAspAlaGluGlyLysLeuArgGluLeu 208
|||||: |||||: |||||: |||||: |||||: |||||: |||||: |||||
2833 .....CTG 2835
209 HisLeuLysSerIleGluValIleGluValProSerIleProGluArg 225
|||||: |||||: |||||: |||||: |||||: |||||: |||||: |||||
2836 CGTGGGGTCTGACGCCCTAAATACATTGATATTCGGAACAGTGGCCAA 2885

```


[illegible]